

**ANALISIS SENTIMEN MASYARAKAT TERHADAP *E-COMMERCE*
PADA MEDIA SOSIAL MENGGUNAKAN METODE
NAÏVE BAYES CLASSIFIER (NBC) DENGAN
SELEKSI FITUR *INFORMATION GAIN (IG)***

SKRIPSI

Oleh :
ABDAN SYAKURO
NIM. 13650129



**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2017**

**ANALISIS SENTIMEN MASYARAKAT TERHADAP *E-COMMERCE*
PADA MEDIA SOSIAL MENGGUNAKAN METODE
NAÏVE BAYES CLASSIFIER (NBC) DENGAN
SELEKSI FITUR *INFORMATION GAIN (IG)***

SKRIPSI

**Diajukan kepada :
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**Oleh :
ABDAN SYAKURO
NIM. 13650129**

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2017**

LEMBAR PERSETUJUAN

**ANALISIS SENTIMEN MASYARAKAT TERHADAP *E-COMMERCE*
PADA MEDIA SOSIAL MENGGUNAKAN METODE
NAÏVE BAYES CLASSIFIER (NBC) DENGAN
SELEKSI FITUR *INFORMATION GAIN (IG)***

SKRIPSI

Oleh :
Abdan Syakuro
NIM. 13650129

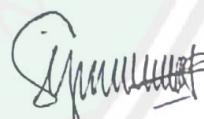
Telah Diperiksa dan Disetujui untuk Diuji
Tanggal, 5 Oktober 2017

Dosen Pembimbing I



Fachrul Kurniawan, M.MT
NIP.19771020 200901 1 001

Dosen Pembimbing II



A'la Syauqi, M.Kom
NIP. 19771201 200801 1 001

Mengetahui,
Ketua Jurusan Teknik Informatika



Dr. Zahyo Crysdian
NIP. 19740424 200901 1 008

LEMBAR PENGESAHAN

ANALISIS SENTIMEN MASYARAKAT TERHADAP *E-COMMERCE* PADA MEDIA SOSIAL MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER (NBC)* DENGAN SELEKSI FITUR *INFORMATION GAIN (IG)*

SKRIPSI

Oleh :




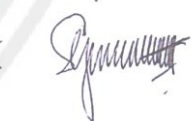
Abdan Syakuro
NIM. 13650129

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal : 28 November 2017

Susunan Dewan Penguji

1. Penguji Utama : Dr. Suhartono, M.Kom
NIP. 19680519 200312 1 001
2. Ketua Penguji : Hani Nurhayati, M.T
NIP. 19780625 200801 2 006
3. Sekretaris Penguji : Fachrul Kurniawan, M.MT
NIP.19771020 200901 1 001
4. Anggota Penguji : A'la Syauqi, M.Kom
NIP. 19771201 200801 1 001

Tanda Tangan

()
()
()
()

Mengesahkan,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang


Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Abdan Syakuro

NIM : 13650129

Jurusan : Teknik Informatika

Fakultas/ Jurusan : Sains dan Teknologi

Judul Skripsi : Analisis Sentimen Masyarakat terhadap *E-Commerce* pada Media Sosial menggunakan Metode *Naïve Bayes Classifier (NBC)* dengan Seleksi Fitur *Information Gain (IG)*.

Menyatakan dengan sebenar-benarnya bahwa hasil penelitian saya ini tidak terdapat unsur-unsur penjiplakan karya penelitian atau karya ilmiah yang pernah dilakukan atau dibuat oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata hasil penelitian ini terbukti terdapat unsur-unsur penjiplakan, maka saya bersedia untuk mempertanggungjawabkan, serta diproses sesuai peraturan yang berlaku.

Malang, 5 Oktober 2017

Yang membuat pernyataan



Abdan Syakuro

NIM. 13650129

MOTTO

“Hidup Mulia atau Mati Syahid”

“Barangsiapa bertakwa kepada Allah niscaya Dia akan mengadakan baginya jalan keluar. Dan memberinya rezki dari arah yang tiada disangka – sangkanya.”

(QS. Ath Tholaq: 2-3)



HALAMAN PERSEMBAHAN

Alhamdulillah, segala puji bagi Allah, Rabb semesta alam yang telah memberikan segala kenikmatan yang tidak bisa hitung jumlahnya. Shalawat serta salam semoga senantiasa ditujukan bagi Rasulullah ﷺ keluarga, para sahabat, dan siapa saja yang meneladani mereka dengan baik hingga hari kiamat.

Terimakasih kepada orang tua saya tercinta, Basuki (Alm) dan Markamah yang telah mendidik saya, mendukung saya dalam melangkah, menemani saya setiap saat, dan setiap hari mendo'akan saya hingga saya mampu untuk menyelesaikan segala kewajiban saya di bangku pendidikan. Kepada dua kakak saya, Aziz Kurniawan dan Ahsan Hakim, terimakasih atas segala dukungan dan motivasinya, semoga keluarga kita kelak dikumpulkan didalam surga – Nya.

Terimakasih kepada para ustadz – ustadz yang telah membimbing saya untuk menjadi hamba Allah yang seutuhnya. Semoga Allah selalu memberikan kesehatan untuk selalu berdakwah, dan mendapatkan derajat yang tinggi di akhirat.

Terimakasih kepada dosen – dosen yang telah sabar dan ikhlas dalam mendidik saya hingga mampu melewati seluruh ujian mata kuliah, terutama kepada Bapak Fachrul Kurniawan, M.MT dan Bapak A'la Syauqi, M.Kom semoga ilmu yang beliau berikan bermanfaat bagi seluruh mahasiswa khususnya Teknik Informatika angkatan 2013 UIN Maliki Malang.

Terimakasih kepada keluargaku di Pesma Al – Ihsan Malang, semoga kelak di akhirat kita termasuk golongan orang yang mendapat naungan Allah dimana tidak ada naungan kecuali naungan – Nya.

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Segala puji bagi Allah, Rabb semesta alam. Shalawat dan salam semoga senantiasa ditujukan bagi Rasulullah SAW, keluarga, para sahabat, dan siapa saja yang meneladani mereka dengan baik hingga hari kiamat.

Dalam menyelesaikan skripsi ini, banyak pihak yang telah memberikan bantuan baik secara moril, nasihat, dan semangat maupun materil. Atas segala bantuan yang telah diberikan, penulis ingin menyampaikan doa dan ucapan terimakasih yang sedalam – dalamnya kepada :

1. Bapak Prof. Dr. Abdul Haris, M.Ag selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Ibu Dr. Sri Harini, M.Si selaku dekan Fakultas Sains dan Teknologi Universitas Maulana Malik Ibrahim Malang beserta seluruh staf.
3. Bapak Dr. Cahyo Crysdiyan, selaku ketua jurusan teknik informatika yang telah memberikan motivasi untuk terus berjuang.
4. Bapak Bapak Fachrul Kurniawan M.MT dan Bapak A'la Syauqi, M.Kom selaku dosen pembimbing penulis yang telah meluangkan waktu untuk membimbing, mengarahkan dan memberi masukan kepada penulis dalam pengerjaan skripsi ini hingga akhir.
5. Seluruh Dosen, Laboran, dan Staff Administrasi Jurusan Teknik Informatika UIN Malang, terima kasih atas segala ilmu dan bimbingannya.
6. Segenap dosen teknik informatika yang telah memberikan bimbingan keilmuan kepada penulis selama masa studi.

7. Seluruh rekan-rekan studi yang tidak dapat disebutkan satu persatu, terima kasih atas segala kebaikan yang diberikan kepada penulis.

Berbagai kekurangan dan kesalahan mungkin pembaca temukan dalam penulisan skripsi ini, untuk itu penulis menerima segala kritik dan saran yang membangun dari pembaca sekalian. Semoga apa yang menjadi kekurangan bisa disempurnakan oleh peneliti selanjutnya dan semoga karya ini senantiasa dapat memberi manfaat.

Wassalamualaikum Wr.Wb.

Malang, 5 Oktober 2017

Penulis



DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN KEASLIAN TULISAN	iv
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
ABSTRAK	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu.....	6
2.2 <i>Data Mining</i>	8
2.3 <i>Text Mining</i>	8
2.4 Analisis Sentimen.....	9
2.5 Media Sosial	10
2.6 <i>Text Preprocessing</i>	10
2.7 Seleksi <i>Fitur</i>	11
2.8 <i>Information Gain (IG)</i>	12
2.9 <i>Naïve Bayes Classifier (NB)</i>	13
2.10 Evaluasi	15
BAB III METODOLOGI PENELITIAN.....	17
3.1 Deskripsi Umum.....	17

3.2	<i>Input</i>	18
3.3	<i>Dataset</i>	18
3.4	<i>Preprocessing</i>	20
3.4.1	<i>Case Folding</i>	21
3.4.2	<i>Cleansing</i>	22
3.4.3	<i>Convert Emoticon</i>	23
3.4.4	<i>Convert Negation</i>	24
3.4.5	<i>Tokenizing</i>	25
3.4.6	<i>Normalization</i>	26
3.4.7	<i>Stopword Removal</i>	28
3.4.8	<i>Stemming</i>	29
3.4.9	<i>Seleksi Fitur Information Gain</i>	31
3.5	<i>Naïve Bayes Classification</i>	34
3.5.1	<i>Proses Training</i>	35
3.5.2	<i>Proses Testing</i>	36
3.6	<i>Perancangan Desain Sistem</i>	39
3.7	<i>Perancangan Antar Muka</i>	40
BAB IV HASIL DAN PEMBAHASAN.....		44
4.1	<i>Implementasi</i>	44
4.1.1	<i>Pengumpulan Data Training</i>	44
4.1.2	<i>Pengumpulan Data Testing</i>	47
4.1.3	<i>Proses Preprocessing Dokumen</i>	48
4.1.4	<i>Pembobotan Information Gain</i>	54
4.1.5	<i>Pembobotan Naïve Bayes</i>	55
4.1.6	<i>Klasifikasi Metode Naïve Bayes</i>	58
4.1.7	<i>Desain dan Implementasi GUI</i>	59
4.2	<i>Pengujian dan Pembahasan Sistem</i>	65
4.3	<i>Integrasi Penelitian dengan Islam</i>	67
BAB V PENUTUP.....		71
5.1	<i>Kesimpulan</i>	71
5.2	<i>Saran</i>	71
DAFTAR PUSTAKA.....		72

DAFTAR GAMBAR

Gambar 3.1 Gambaran Umum Sistem	17
Gambar 3.2 Diagram Proses <i>Dataset</i>	19
Gambar 3.3 Diagram Alir <i>Preprocessing</i>	20
Gambar 3.4 Diagram Alir <i>Case Folding</i>	21
Gambar 3.5 Diagram Alir <i>Cleansing</i>	22
Gambar 3.6 Diagram Alir <i>Convert Emoticon</i>	23
Gambar 3.7 Diagram Alir <i>Convert Negation</i>	24
Gambar 3.8 Diagram Alir <i>Tokenizing</i>	25
Gambar 3.9 Diagram Alir <i>Normalization</i>	26
Gambar 3.10 Diagram Alir <i>Stopword Removal</i>	28
Gambar 3.11 Diagram Alir <i>Stemming</i>	31
Gambar 3.12 Diagram Alir Metode <i>Information Gain</i>	32
Gambar 3.13 Diagram Alir Metode <i>Naïve Bayes</i>	34
Gambar 3.14 Desain Sistem	39
Gambar 3.15 Tampilan Halaman Utama	40
Gambar 3.16 Tampilan <i>Crawling</i>	41
Gambar 3.17 Tampilan <i>Preprocessing</i>	41
Gambar 3.18 Tampilan Bobot <i>Information Gain</i>	42
Gambar 3.19 Tampilan Bobot <i>Naïve Bayes</i>	42
Gambar 3.20 Tampilan Klasifikasi <i>Naïve Bayes</i>	43
Gambar 3.21 Tampilan Visualisasi	43
Gambar 4.1 Pencarian Kata Kunci	45
Gambar 4.2 <i>Fitur Inspect Element</i>	46
Gambar 4.3 Fungsi Koneksi dengan Token	47
Gambar 4.4 Fungsi <i>Crawling Data Testing</i>	48
Gambar 4.5 Fungsi Proses <i>Case Folding</i>	49
Gambar 4.6 Fungsi Proses <i>Cleansing</i>	50

Gambar 4.7 Fungsi Convert Emoticon.....	51
Gambar 4.8 Fungsi Convert Negation	52
Gambar 4.9 Fungsi Tokenizing.....	52
Gambar 4.10 Fungsi Normalization.....	53
Gambar 4.11 Fungsi Stopword Removal.....	54
Gambar 4.12 Fungsi Stemming.....	54
Gambar 4.13 Fungsi Information Gain	55
Gambar 4.14 Fungsi Probabilitas Kata Positif.....	56
Gambar 4.15 Fungsi Probabilitas Kata Negatif	57
Gambar 4.16 Fungsi Naïve Bayes Classifier	59
Gambar 4.17 Tampilan Halaman Utama	60
Gambar 4.18 Tampilan Proses Crawling	61
Gambar 4.19 Tampilan Proses <i>Preprocessing</i>	61
Gambar 4.20 Tampilan Bobot <i>Information Gain</i>	62
Gambar 4.21 Tampilan Proses Bobot Naïve Bayes	63
Gambar 4.22 Tampilan Proses Klasifikasi Sentimen.....	63
Gambar 4.23 Tampilan Visualisasi.....	64
Gambar 4.24 Grafik Akurasi.....	67

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	16
Tabel 3.1 Dokumen <i>Tweet</i>	18
Tabel 3.2 Contoh Tahap <i>Case Folding</i>	21
Tabel 3.3 Contoh Tahap <i>Cleansing</i>	23
Tabel 3.4 Contoh Tahap <i>Convert Emoticon</i>	24
Tabel 3.5 Contoh Tahap <i>Convert Negation</i>	25
Tabel 3.6 Contoh Tahap <i>Tokenizing</i>	26
Tabel 3.7 Contoh Tahap Konversi Kata Singkatan.....	27
Tabel 3.8 Contoh Tahap Konversi Kata Baku	27
Tabel 3.9 Contoh Tahap Konversi Kata Inggris	28
Tabel 3.10 Contoh Tahap <i>Stopword Removal</i>	29
Tabel 3.11 Contoh Tahap <i>Stemming</i>	31
Tabel 3.12 Contoh Koleksi Data.....	32
Tabel 3.13 Contoh Kasus Data <i>Training</i>	35
Tabel 3.14 Perhitungan Probabilitas Data <i>Training</i>	35
Tabel 3.15 Contoh Kasus Data Testing	36
Tabel 3.16 Nilai Probabilitas Data Testing.....	38
Tabel 4.1 Rincian Jumlah Data <i>Training</i>	46
Tabel 4.2 Hasil Akurasi.....	65

ABSTRAK

Syakuro, Abdan. 2017. **Analisis Sentimen Masyarakat terhadap E-Commerce pada Media Sosial menggunakan Metode Naïve Bayes Classifier (NBC) dengan Seleksi Fitur Information Gain (IG)**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing : (1) Fachrul Kurniawan M.MT (II) A'la Syauqi, M.Kom

Kata Kunci : Analisis Sentimen, Media Sosial, Klasifikasi, *Naive Bayes Classifier*, *Information Gain*.

Media sosial pada saat ini merupakan media komunikasi yang sangat populer dikalangan masyarakat Indonesia. Salah satu media sosial yang sedang digandrungi masyarakat saat ini adalah *twitter*. Dengan media sosial *twitter*, informasi masyarakat sangat beragam, dari informasi tersebut terdapat data yang dapat diolah menjadi analisa sentimen. Penelitian ini bertujuan untuk membuat sistem analisis sentimen masyarakat yang berupa sentimen positif atau negatif. Metode yang digunakan untuk klasifikasi sentimen ini adalah *Naïve Bayes Classifier* dengan Seleksi Fitur *Information Gain*. Masukan sistem ini berupa data *tweet* dari *e-commerce* akun Lazada, Tokopedia, dan Bukalapak, sedangkan hasil keluaran sistem ini berupa visualisasi data sentimen positif dan negatif. Pembuatan sistem analisis sentimen menggunakan metode *Naïve Bayes* dengan seleksi fitur *Information Gain* terbukti bahwa algoritma tersebut dapat menganalisis sentimen secara otomatis, dengan tingkat akurasi 88,8 %. Hasil visualisasi ini dapat digunakan oleh pihak *e-commerce* untuk analisa produknya untuk menentukan kebijakan yang diambil pada masa depan.

ABSTRACT

Syakuro, Abdan. 2017. **The Sentiment Analysis of Society towards E-Commerce on Social Media uses the Naive Bayes Classifier (NBC) Method with Information Gain Feature Selection (IG).** Undergraduate Thesis. Informatics Engineering Departement. Faculty of Science and Technology. State Islamic University of Maulana Malik Ibrahim Malang.

Advisor : (I) Fachrul Kurniawan, M.MT (II) A'la Syauqi, M.Kom

Keywords : Sentiment Analysis, Social Media, Classification, Naive Bayes Classifier, Information Gain.

Social media at this time is a very popular communication media among the people of Indonesia. One of the social media that is being loved today is twitter. With social media twitter, public information is very diverse, from the information there is data that can be processed into a sentiment analysis. This study aims to create a community sentiment analysis system in the form of positive or negative sentiments. The method used for the classification of these sentiments is the Naïve Bayes Classifier with Information Gain Features Selection. Input this system in the form of tweet data from the e-commerce account Lazada, Tokopedia, and Bukalapak, while the output of this system in the form of visualization of positive and negative sentiment data. Making a system of sentiment analysis using Naïve Bayes method with Information Gain feature selection proven that the algorithm can analyze sentiment automatically, with 88.8% accuracy level. This visualization result can be used by the e-commerce to analyze its products to determine the policy taken in the future.

الملخص

شكورا، عبدان. 2017. تحليل معنويات المجتمع للتجارة الإلكترونية على وسائل الاعلام الاجتماعية باستخدام الطريقة (NBC) **Information Gain (IG) & Naïve Bayes Classifier**. البحث العلمي. قسم المعلوماتية كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالنج. تحت الإشراف : فخرول كرنبوا الماجستير، أعلى شوق الماجستير.

كلمت البحث : تحليل المشاعر، وسائل الاعلام الاجتماعية، التصنيف، Information Gain, Naive Bayes

وسائل الإعلام الاجتماعية في هذا الوقت هو وسائل الإعلام شعبية جدا بين شعب اندونيسيا. واحدة من وسائل الاعلام الاجتماعية التي يجري محبوبا اليوم هو تويتر. مع تويتر وسائل الاعلام الاجتماعية، والمعلومات العامة هي متنوعة جدا، من المعلومات هناك البيانات التي يمكن معالجتها في تحليل المشاعر. تهدف هذه الدراسة إلى إنشاء نظام تحليل المشاعر المجتمعية في شكل مشاعر إيجابية أو سلبية. الطريقة المستخدمة لتصنيف المشاعر هذه هي Naïve Bayes Classifier & Information Gain. يتم إدخال هذا الإدخال في النظام من حساب البريد الإلكتروني البالغ عددهم Lazada, Tokopedia, Bukalapak حسابا، في حين أن إنتاج هذا النظام في شكل تصور بيانات المشاعر الإيجابية والسلبية. جعل نظام تحليل المشاعر باستخدام طريقة Naïve Bayes & Information Gain أثبتت أن الخوارزمية يمكن تحليل المشاعر تلقائيا، مع دقة 88,8%. هذه النتيجة التصور يمكن استخدامها من قبل التجارة الإلكترونية لتحليل منتجاتها لتحديد السياسة المتخذة في المستقبل.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring berkembangnya teknologi internet di Indonesia, memunculkan banyak situs jual beli *online* atau yang lebih populer disebut *e-commerce*. Saat ini *e-commerce* merupakan tempat jual beli *online* yang semakin diminati oleh masyarakat kota. Hal ini dikarenakan kemudahan transaksi tanpa harus datang ke toko fisik, dan setiap tahunnya jumlah pengguna *e-commerce* semakin meningkat. Menurut data yang dirilis biro riset Frost & Sullivan, bersama China, Indonesia menjadi negara dengan pertumbuhan pasar *e-commerce* terbesar di dunia dengan rata – rata pertumbuhan 17 persen setiap tahun (tekno.liputan6.com, 2015).

Dari peningkatan jumlah pengguna *e-commerce* tersebut, tentunya kejahatan atau penipuan dalam dunia maya juga semakin meningkat. Dalam riset yang dilakukan oleh Kaspersky Lab dan B2B Internasioanl, terungkap bahwa 48% konsumen menjadi target aksi penipuan yang dirancang untuk menipu dan mengelabui mereka sehingga mengungkapkan informasi sensitif dan data keuangan untuk tindak kriminal. Hal yang mengkhawatirkan, dari 26 negara yang disurvei, Indonesia menempati posisi tertinggi sebesar 26% konsumen telah kehilangan uang mereka sebagai akibat menjadi target aksi penipuan *online* (inet.detik.com, 2016). Dari banyaknya *e-commerce* yang ada di Indonesia maka pasti semakin banyak kejahatan *online* yang terjadi, sehingga penilaian masyarakat terhadap suatu *e-commerce* dapat dijadikan analisa terhadap pasar *online*.

Untuk mengatasi masalah tersebut, opini masyarakat terhadap suatu *e-commerce* dapat membantu masyarakat lain supaya lebih berhati – hati dalam melakukan transaksi *online*. Penulis merasa perlu untuk melakukan penelitian ini yaitu dengan membuat sebuah sistem penganalisa opini atau biasa disebut analisis sentimen masyarakat sehingga bisa mengetahui dan membantu memberikan informasi mengenai analisa sentimen *e-commerce* masyarakat.

Adapun *e-commerce* yang menjadi objek penelitian ini adalah Lazada, Bukalapak, dan Tokopedia. Dilansir dari artikel yang ditulis oleh media *online* tekno.liputan6.com, ketiga *e-commerce* tersebut termasuk lima situs *e-commerce* terbaik di Indonesia yang paling sering dikunjungi konsumen. Data tersebut diperoleh dari situs Alexa dengan peringkat pertama adalah Bukalapak, yang kedua adalah Lazada dan yang ketiga adalah Tokopedia.

Opini masyarakat dapat diperoleh dari berbagai media cetak maupun elektronik. Masyarakat kota saat ini lebih sering menggunakan media sosial dalam mengomentari suatu masalah termasuk suatu produk. Salah satu media sosial yang digemari masyarakat Indonesia saat ini adalah *Twitter*.

Orang Indonesia dikenal sangat aktif di media sosial. *Country Business Head Twitter* Indonesia Roy Simangunsong mengatakan, jumlah cuitan orang Indonesia selama Januari hingga Desember 2016 mencapai 4,1 miliar *tweet*. Meski tak menyebut jumlah pengguna di Indonesia, Roy mengatakan bahwa jumlah pengguna aktif *Twitter* Indonesia mencapai 77 persen dari seluruh pengguna di dunia (tekno.liputan6.com, 2016). Tentu saja, informasi yang terkandung dalam *tweet* ini sangat berharga sebagai alat penentu kebijakan dan ini bisa dilakukan dengan *Text Mining*.

Text Mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dokumen dimana *Text Mining* merupakan variasi dari *Data Mining* yang berusaha menemukan pola menarik dari sekumpulan data tekstual yang berjumlah besar (Kurniawan, Effendi, & Sitompul, 2012).

Analisis Sentimen atau *Opinion Mining* adalah studi komputasional dari opini – opini orang, sentimen dan emosi melalui entitas atau atribut yang dimiliki yang diekspresikan dalam bentuk teks (Aditya, Hani'ah, Fitrawan, Arifin, & Purwitasari, 2016). Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif, atau netral (Pang & Lee, Opinion mining and sentiment analysis, 2008).

Masalah analisis sentimen sangat penting untuk diperhatikan oleh seorang muslim karena Allah berfirman dalam Al-Qur'an surah Al - Hujurat ayat 6 yang artinya:

"Hai orang-orang yang beriman, jika datang kepadamu orang fasik, membawa suatu berita, maka periksalah dengan teliti, agar kamu tidak menimpakan suatu musibah, kepada suatu kaum, tanpa mengetahui keadaannya, yang menyebabkan kamu menyesal atas perbuatanmu itu."

Ayat tersebut menerangkan bahwa jika seseorang mendapatkan opini masyarakat yang belum jelas, seharusnya melakukan *tabayyun* dahulu. Karena jika tidak melakukan *tabayyun* dapat mengakibatkan dampak buruk bagi diri sendiri maupun orang lain. Berangkat dari ayat tersebut, maka penulis membangun sebuah alat yang salah satu fungsinya adalah *tabayyun* dengan

mengumpulkan berbagai opini dari masyarakat lalu menganalisisnya sehingga mendapatkan hasil yang berguna bagi masyarakat.

Berdasarkan penelitian (Chandani, Wahono, & Purwanto, 2015) yang meneliti perbandingan akurasi antara *information gain*, *chi square*, *forward selection*, *backward selection* didapatkan *information gain* merupakan yang terbaik. Sehingga pada penelitian ini penulis akan membuat aplikasi Analisis Sentimen menggunakan metode *Naïve Bayes* untuk klasifikasi dengan seleksi *fitur information gain*.

1.2 Rumusan Masalah

- 1) Apakah metode *Naïve Bayes Classifier* dengan seleksi *fitur Information Gain* dapat digunakan dalam pengklasifikasian Analisis Sentimen *E-Commerce*.
- 2) Seberapa baik performa metode *Naïve Bayes Classifier* dengan seleksi *fitur Information Gain* dalam pengklasifikasian Analisis Sentimen *E-Commerce*?

1.3 Tujuan Penelitian

- 1) Membuktikan bahwa metode *Naïve Bayes Classifier* dengan Seleksi *Fitur Information Gain* dapat digunakan dalam pengklasifikasian Analisis Sentimen *E-Commerce*.
- 2) Mengukur efisiensi metode *Naive Bayes Classifier* dengan Seleksi *Fitur Information Gain*.

1.4 Manfaat Penelitian

- 1) Memberikan khazanah keilmuan dan menambah informasi khususnya pertimbangan dalam permasalahan – permasalahan *e-commerce*.

- 2) Memberikan informasi sentimen positif atau negatif kepada pemilik produk agar bisa menganalisa produknya.

1.5 Batasan Masalah

- 1) Data yang dianalisis adalah data dari media sosial berbahasa Indonesia.
- 2) Metode yang digunakan untuk klasifikasi adalah *Naïve Bayes* dengan seleksi fitur *Information Gain*.
- 3) Data yang dijadikan data *testing* hanya 7 hari ke belakang dari hari proses *crawling*.
- 4) Jumlah data *training* yang digunakan adalah 3000 *tweet*.

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian mengenai klasifikasi sentimen telah dilakukan oleh Bo Pang(2002). Pada jurnalnya, Bo Pang melakukan klasifikasi sentimen terhadap *review* film dengan menggunakan berbagai teknik pembelajaran mesin. Teknik pembelajaran mesin yang digunakan yaitu *Naïve Bayes*, *Maximum Entropy*, dan *Support Vector Machines (SVM)*. Pada penelitian itu juga digunakan beberapa pendekatan untuk melakukan ekstraksi fitur, yaitu *unigram*, *unigram + bigram*, *unigram + Part of Speech (POS)*, *adjective*, dan *unigram + posisi*. Hasil dari eksperimen yang dilakukan dipenelitian ini menemukan bahwa *SVM* menjadi metode terbaik ketika dikombinasikan dengan *unigram* dengan akurasi 82.9% (Pang & Lee, Thumbs up? Sentiment Classification using Machine Learning, 2002).

Pada penelitian yang dilakukan oleh Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2012) dikembangkan sistem *opinion mining* untuk menganalisis opini publik pada perguruan tinggi. Pada subproses *document subjectivity* dan *target detection* digunakan *Part-of-Speech (POS) Tagging* menggunakan *Hidden Markov Model (HMM)*. Pada hasil proses *POSTagging* kemudian diterapkan *rule* untuk mengetahui apakah suatu dokumen termasuk opini atau bukan, serta untuk mengetahui bagian kalimat mana yang merupakan objek yang menjadi target opini. Dokumen yang dikenali sebagai opini selanjutnya diklasifikasikan ke dalam opini negatif dan positif (*subproses opinion orientation*) menggunakan *Naïve*

Bayes Classifier (NBC). Dari pengujian didapatkan nilai *precision* dan *recall* untuk subproses *document subjectivity* adalah 0.99 dan 0.88, untuk subproses target *detection* adalah 0.92 dan 0.93, serta untuk subproses *opinion orientation* adalah 0.95 dan 0.94. (Rozi, Pranomo, & Dahlan, 2012)

Penelitian tentang klasifikasi juga dilakukan oleh Rodiyansyah, S. F. dan Winarko Edi dengan melakukan teknik *Data Mining* yang digunakan untuk visualisasi kemacetan lalu lintas di sebuah kota. Pada penelitiannya ini metode yang dipakai adalah *Naïve Bayes* dengan mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Dari hasil uji coba, aplikasi menunjukkan bahwa nilai akurasi terkecil 78% dihasilkan pada pengujian dengan sampel sebanyak 100 dan menghasilkan nilai akurasi tinggi 91,60% pada pengujian dengan sampel sebanyak 13106. Hasil pengujian dengan perangkat lunak *Rapid Miner 5.1* diperoleh nilai akurasi terkecil 72% dengan sampel sebanyak 100 dan nilai akurasi tertinggi 93,58% dengan sampel 13106 untuk metode *Naive Bayesian Classification*. Sedangkan untuk metode *Support Vector Machine* diperoleh nilai akurasi terkecil 92% dengan sampel sebanyak 100 dan nilai akurasi tertinggi 99,11% dengan sampel sebanyak 13106. (Rodiyansyah & Winarko, 2013)

Penelitian lainnya pada klasifikasi sentimen *review* film dilakukan oleh Chandani, V., Wahono, R. S., & Purwanto (2015) dengan mengkomparasi metode klasifikasi *Machine Learning* seperti *Naïve Bayes(NB)*, *Support Vector Machine (SVM)*, dan *Artificial Neural Network (ANN)* dan seleksi fitur seperti *Information Gain*, *Chi Square*, *Forward Selection* dan *Backward Elimination*. Hasil komparasi metode *SVM* mendapatkan hasil yang terbaik dengan akurasi 81.10% dan *AUC*

0.904. Hasil dari komparasi seleksi *Fitur Information Gain* mendapatkan hasil yang paling baik dengan rata – rata akurasi 84.57% dan rata – rata *AUC* 0.899. Hasil integrasi metode klasifikasi terbaik dan metode seleksi *fitur* terbaik menghasilkan akurasi 81.50% dan *AUC* 0.929. Hasil ini mengalami kenaikan jika dibandingkan hasil eksperimen yang menggunakan *SVM* tanpa seleksi *fitur*. Hasil dari pengujian metode seleksi *fitur* terbaik adalah *Information Gain* mendapatkan hasil terbaik untuk digunakan pada metode *NB*, *SVM* dan *ANN* (Chandani, Wahono, & Purwanto, 2015).

2.2 *Data Mining*

Data Mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. *Data Mining* merupakan proses *iteratif* dan interaktif untuk mengemukakan pola atau model baru sah (sempurna), bermanfaat dan dapat dimengerti dalam suatu *database* yang sangat besar (*massive database*).

Data Mining berisi pencarian tren atau pola yang diinginkan dalam *database* besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola – pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya (Hermawati, 2013).

2.3 *Text Mining*

Text mining adalah proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari

kata – kata yang dapat mewakili isi dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen tersebut. (Aditya B. R., 2015)

Text mining juga dikenal sebagai *data mining text* atau penemuan pengetahuan dari *database* tekstual. Sesuai dengan buku *The Text Mining Handbook*, *text mining* dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen – komponen dalam *data mining*. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan dalam *text mining* adalah sekumpulan teks yang memiliki format yang tidak terstruktur atau minimal *semi* terstruktur. Adapun tugas khusus dari *text mining* antara lain yaitu pengkategorisasian teks dan pengelompokkan teks. (Nurhuda, Sihwi, & Doewes, 2013)

2.4 Analisis Sentimen

Menurut Medhat et al (2014, 1093) analisis sentimen adalah suatu bidang yang berlangsung dalam penelitian berbasis teks. Analisis sentimen atau *opinion mining* adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas, dimana entitas tersebut dapat mewakili individu. (Wati, 2016)

Analisis Sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah

masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif. (Rozi, Pranomo, & Dahlan, 2012)

2.5 Media Sosial

Pada dasarnya media sosial merupakan perkembangan mutakhir dari teknologi – teknologi web baru berbasis internet, yang memudahkan semua orang untuk dapat berkomunikasi, berpartisipasi, saling berbagi dan membentuk sebuah jaringan secara *online*, sehingga dapat menyebarluaskan konten mereka sendiri. Seperti *post* di blog, *tweet*, atau video *youtube* dapat direproduksi dan dapat dilihat secara langsung oleh jutaan orang secara gratis.

Media sosial mempunyai banyak bentuk, diantaranya yang paling populer yaitu *microblogging* (*twitter*), facebook, dan blog. *Twitter* adalah suatu situs web yang merupakan layanan dari *microblog*, yaitu suatu bentuk blog yang membatasi ukuran setiap posnya, yang memberikan fasilitas bagi pengguna untuk dapat menuliskan pesan dalam *twitter update* hanya berisi 140 karakter. *Twitter* merupakan salah satu jejaring sosial yang paling mudah digunakan, karena hanya memerlukan waktu yang singkat tetapi informasi yang disampaikan dapat langsung menyebar secara luas. (Setyani, 2013)

2.6 Text Preprocessing

Preprocessing merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *preprocessing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah yang diproses oleh sistem. *Preprocessing* sangat penting dalam pembuatan *analisis sentimen*, terutama untuk media sosial yang sebagian besar berisi kata – kata atau kalimat yang tidak formal dan tidak terstruktur serta

memiliki *noise* yang besar. Ada tiga model *preprocessing* untuk kalimat atau teks dengan *noise* yang besar (A Clark, 2003). Tiga model tersebut adalah :

1. *Orthographic Model*. Model ini dipergunakan untuk memperbaiki kata atau kalimat yang memiliki kesalahan dari segi bentuk kata atau kalimat. Contoh kesalahan yang diperbaiki dengan *Orthographic model* adalah huruf kapital di tengah kata.
2. *Error Model*. Model ini dipergunakan untuk memperbaiki kesalahan dari segi kesalahan eja atau kesalahan penulisan. Ada dua jenis kesalahan yang dikoreksi dengan model ini yaitu kesalahan penulisan dan kesalahan eja. Kesalahan penulisan mengacu pada kesalahan pengetikan sedangkan kesalahan eja muncul ketika penulis tidak tahu ejaannya benar atau salah.
3. *White Space Model*. Model ke tiga ini mengacu pada pengoreksian tanda baca. Contoh kesalahan untuk model ini adalah tidak menggunakan tanda titik ‘.’ di akhir kalimat. Namun, model ini tidak terlalu signifikan, terutama ketika berhadapan dengan media sosial yang jarang menginginkan tanda baca. (Mujilahwati, 2016)

2.7 Seleksi *Fitur*

Seleksi *fitur* adalah salah satu teknik terpenting dan sering digunakan dalam *preprocessing*. Teknik ini mengurangi jumlah *fitur* yang terlibat dalam menentukan suatu nilai kelas target, mengurangi *fitur irrelevant*, berlebihan dan data yang menyebabkan salah pengertian terhadap kelas target yang membuat efek segera bagi aplikasi. Tujuan utama dari seleksi *fitur* ialah memilih *fitur* terbaik dari suatu kumpulan *fitur* data. (Maulida, Suyatno, & Hatta, 2016)

2.8 Information Gain (IG)

Information Gain merupakan teknik seleksi *fitur* yang memakai metode *scoring* untuk nominal ataupun pembobotan atribut kontinu yang didiskritkan menggunakan maksimal *entropy*. Suatu *entropy* digunakan untuk mendefinisikan nilai Information Gain. *Entropy* menggambarkan banyaknya informasi yang dibutuhkan untuk mengkodekan suatu kelas. Information Gain (IG) dari suatu *term* diukur dengan menghitung jumlah *bit* informasi yang diambil dari prediksi kategori dengan ada atau tidaknya *term* dalam suatu dokumen. (Maulida, Suyatno, & Hatta, 2016)

Teknik seleksi *fitur* dengan *information gain* artinya adalah memilih simpul *fitur* dari pohon keputusan berdasar nilai *information gain*. Nilai *information gain* sebuah *fitur* diukur dari pengaruh *fitur* tersebut terhadap keseragaman kelas pada data yang dipecah menjadi subdata dengan nilai *fitur* tertentu. Keseragaman kelas (*entropy*) dihitung pada data sebelum dipecah dengan persamaan 2.1 dan pada data setelah dipecah dengan persamaan 2.2 berikut ini.

$$Entropy(S) = \sum_{i=1}^k (P_i) \log_2(P_i) \quad (2.1)$$

Dengan nilai P_i adalah proporsi data S dengan kelas i . K adalah jumlah kelas pada *output* S .

$$Entropy(S, A) = \sum_{v=1}^v \left(\frac{S_v}{S} * Entropy(S_v) \right) \quad (2.2)$$

Dengan nilai v adalah semua nilai yang mungkin dari atribut A , S_v adalah subset sari S dimana atribut A bernilai v . Nilai *information gain* dihitung dengan persamaan 2.3 berikut ini:

$$Gain(S, A) = Entropy(S) - Entropy(S, A) \quad (2.3)$$

Dengan nilai Gain (S,A) adalah nilai *information gain*. *Entropy* (S) adalah nilai *entropy* sebelum pemisah. *Entropy*(S,A) adalah nilai *entropy* setelah pemisah. Besarnya nilai *information gain* menunjukkan seberapa besar pengaruh suatu atribut terhadap pengklasifikasian data. (Rasywir & Purwarianti, 2015)

2.9 *Naïve Bayes Classifier* (NB)

Naïve Bayes Classifier adalah salah satu metode yang populer digunakan untuk keperluan *data mining* karena kemudahan penggunaannya (Hall, 2006) serta waktu pemrosesannya yang cepat, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan tingkat efektifitas yang tinggi (Taheri & Mammadov, 2013).

Dengan bahasa yang lebih sederhana, *Naïve Bayes Classifier* mengasumsikan bahwa keberadaan maupun ketidakberadaan sebuah *fitur* dalam sebuah kelas tidak memiliki keterkaitan dengan keberadaan maupun ketidakberadaan *fitur* lainnya. Sebagai contoh, sesuatu yang berwarna merah, bulat, dan memiliki diameter sekitar 10 cm bisa dikategorikan sebagai buah apel. Walaupun *fitur* ini bergantung antara satu *fitur* dengan *fitur* yang lainnya. *Naïve Bayes Classifier* akan tetap menganggap bahwa *fitur* – *fitur* tersebut independen dan tidak memiliki pengaruh satu sama lainnya (Rocha, 2006).

Bergantung pada model *probabilitasnya*, *Naïve Bayes Classifier* dapat dilatih untuk melakukan *supervised learning* dengan sangat efektif. Dalam berbagai macam penerapannya, estimasi parameter untuk model *Naïve Bayes* menggunakan metode *maximum likelihood*, yang artinya pengguna dapat menggunakan model *Naïve Bayes* tanpa perlu mempercayai *probabilitas Bayesian* atau tanpa menggunakan metode *Bayesian* (Hadna, Santosa, & Winarno, 2016).

Teorema *Bayes* merupakan teorema yang mengacu pada konsep *probabilitas* bersyarat. Secara umum teorema *Bayes* dapat dinotasikan pada persamaan 2.4 berikut:

$$P(A|B) = \frac{P(A|B)P(A)}{P(B)} \quad (2.4)$$

Pada *Naive Bayes Classification* setiap *tweet* direpresentasikan dalam pasangan atribut $(a_1, a_2, a_3, \dots, a_n)$ dimana a_1 adalah kata pertama a_2 adalah kata kedua dan seterusnya, sedangkan V adalah himpunan kelas. Pada saat klasifikasi, metode ini akan menghasilkan kategori / kelas yang paling tinggi *probabilitasnya* (V_{MAP}) dengan memasukkan atribut $(a_1, a_2, a_3, \dots, a_n)$. Adapun rumus V_{MAP} dapat dilihat pada persamaan 2.5 berikut:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2, a_3, \dots, a_n) \quad (2.5)$$

Dengan menggunakan teorema Bayes, maka persamaan (2.5) dapat ditulis menjadi,

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (2.6)$$

$P(a_1, a_2, a_3, \dots, a_n)$ nilainya konstan untuk semua v_j sehingga persamaan (2.6) dapat juga dinyatakan menjadi persamaan 2.7 berikut:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j) \quad (2.7)$$

Naive Bayes Classifier menyederhanakan hal ini dengan mengasumsikan bahwa didalam setiap kategori, setiap atribut bebas bersyarat satu sama lain. Dengan kata lain,

$$P(a_1, a_2, a_3, \dots, a_n | V_j) = \prod_i P(a_i | v_j) \quad (2.8)$$

Kemudian apabila persamaan (2.7) disubstitusikan ke persamaan (2.8), maka akan menghasilkan persamaan 2.9 berikut:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \times \prod_i P(a_i | v_j) \quad (2.9)$$

$P(v_j)$ dan *probabilitas* kata a_i untuk setiap kategori $P(a_i | v_j)$ dihitung pada saat *training* yang dirumuskan seperti berikut,

$$P(v_j) = \frac{\text{docs}_j}{\text{training}} \quad (2.10)$$

$$P(a_i | v_j) = \frac{n_i + 1}{n + \text{kosakata}} \quad (2.11)$$

Dimana docs_j adalah jumlah dokumen pada kategori j dan *training* adalah jumlah dokumen yang digunakan dalam proses *training*. Sedangkan n_i adalah jumlah kemunculan kata a_i pada kategori v_j , n adalah jumlah kosakata yang muncul pada kategori v_j dan kosakata adalah jumlah kata unik pada semua data *training*. (Rodiyanayah & Winarko, 2013)

2.10 Evaluasi

Evaluasi performasi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performasi dari sistem yang telah dibuat. Parameter pengujian yang digunakan untuk evaluasi yaitu akurasi yang perhitungannya dari *table confusion matrix* (matriks klasifikasi atau tabel *kontigensi*). Tabel 2.1 menampilkan sebuah *confusion matrix* untuk pengklasifikasian kedalam dua kelas (Novantirani, Sabariah, & Effendy, 2015).

Tabel 2.1 *Confusion Matrix*

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive(TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative(TN)

Matriks tersebut memiliki empat nilai yang dijadikan acuan dalam perhitungan, yaitu :

- a) *True Positive* (TP), ketika kelas yang diprediksi positif dan faktanya positif.
- b) *True Negative* (TN), ketika kelas yang diprediksi negatif, dan faktanya negatif.
- c) *False Positive* (FP), ketika kelas yang diprediksi positif dan faktanya negatif.
- d) *False Negative* (FN), ketika kelas yang diprediksi negatif dan faktanya positif.

Alternatif dalam menilai sebuah sistem adalah dengan akurasi. Akurasi adalah ketepatan suatu sistem melakukan klasifikasi yang benar. Perhitungan untuk akurasi dapat dikalkulasi dengan persamaan 2.12 berikut:

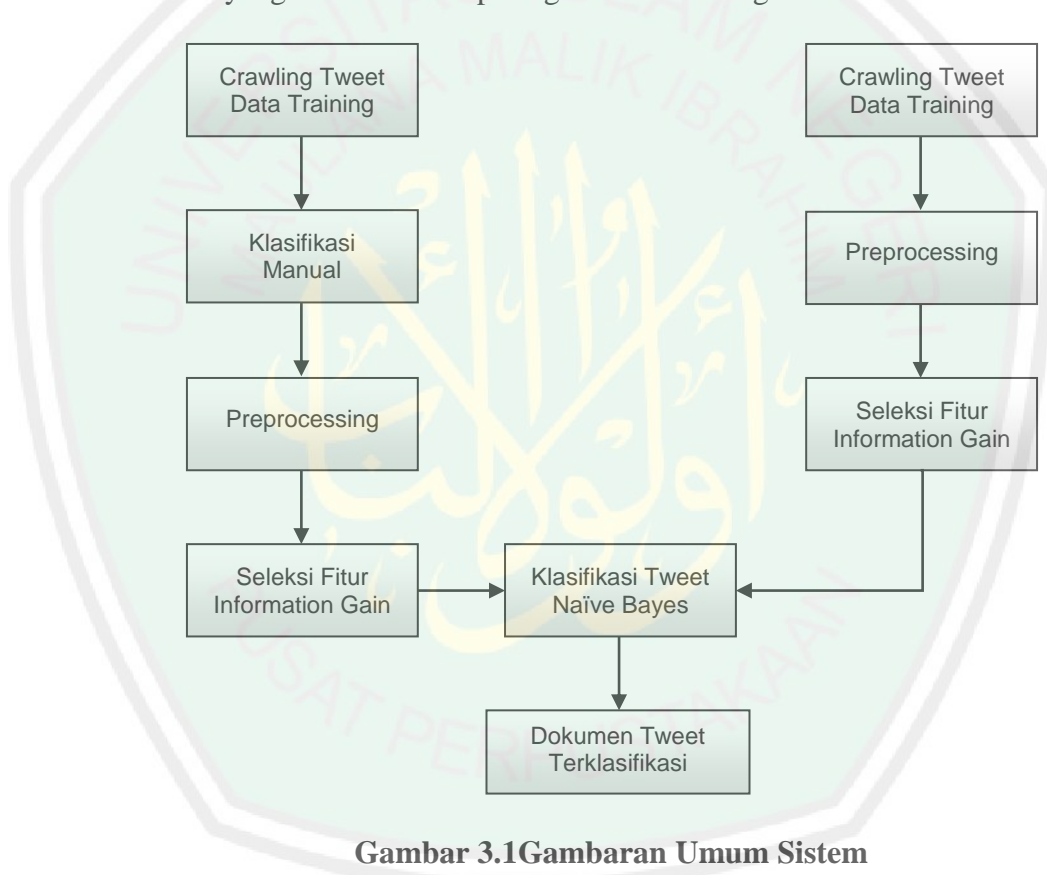
$$Accuracy (A) = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (2.12)$$

BAB III

METODOLOGI PENELITIAN

3.1 Deskripsi Umum

Sistem klasifikasi untuk analisis sentimen yang sedang dikerjakan ini memiliki suatu rancangan bagaimana alur sistem ini akan berjalan. Gambaran umum sistem yang akan dibuat seperti gambar 3.1 sebagai berikut.



Gambar 3.1 Gambaran Umum Sistem

Sistem ini dimulai dengan proses *input* data *tweet* dengan cara *crawling*. Proses *crawling* dibedakan menjadi dua, yaitu *crawling* data *training* dengan cara manual menggunakan *web browser* dan fitur *inspect element*, dan *crawling* data *testing* otomatis menggunakan *API twitter*. Output dari sistem ini merupakan data *testing* dengan nilai *output* berupa opini positif atau negatif yang diklasifikasikan oleh sistem berdasarkan pembelajaran data *training*.

3.2 Input

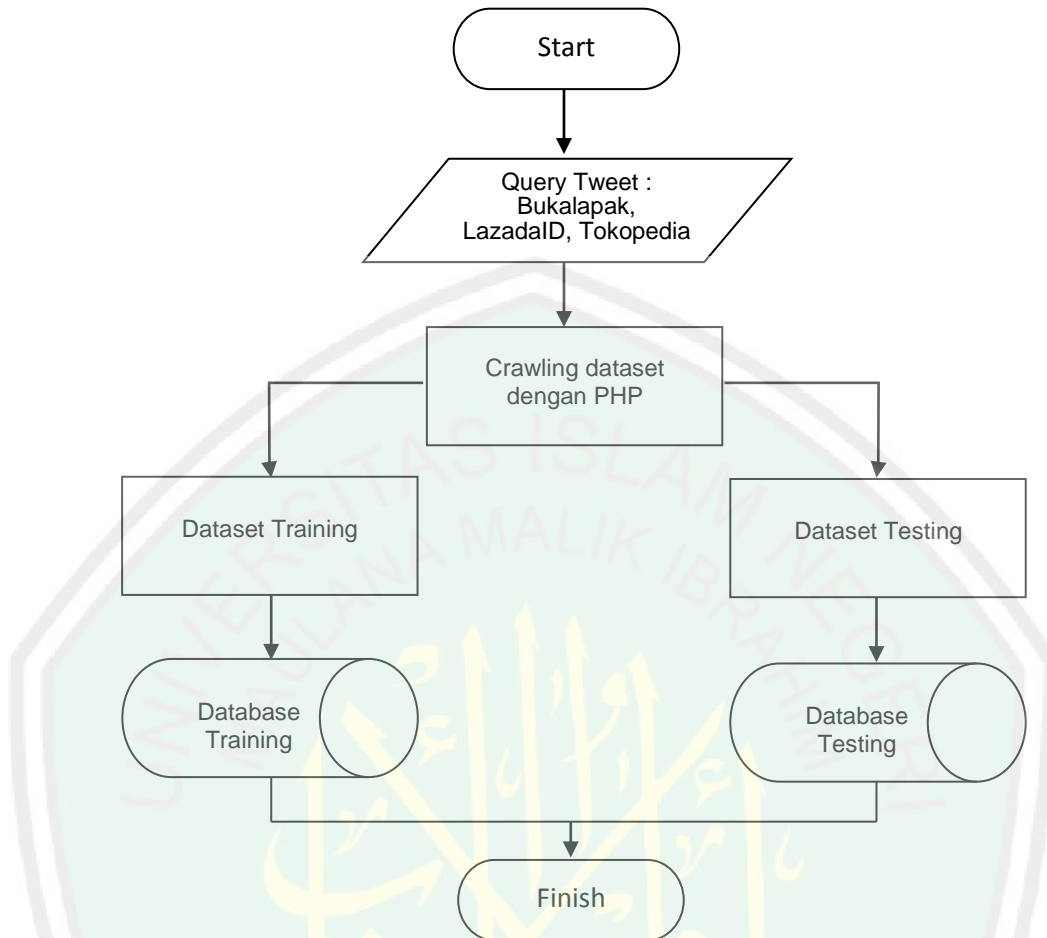
Input yang dimasukkan sistem adalah dokumen berupa *tweet* dari akun *Twitter* yang berupa opini. Data *tweet* tersebut didapat dengan memanfaatkan *fitur API (Application Interface)* yang telah disediakan oleh *Twitter*. Dokumen yang dimasukkan merupakan dokumen berbahasa Indonesia, seperti yang ditunjukkan pada table 3.1 berikut.

Tabel 3.1 Dokumen *Tweet*

Tweet	Opini
	P
	N

3.3 Dataset

Dataset berupa teks berbahasa Indonesia yang diambil dari *website* <http://www.twtiiter.com>. Data yang diambil untuk penelitian ini data yang diambil menggunakan *query* 'bukalapak', 'bukalapak_care', 'lazadaid', 'lazadaidcare', 'tokopedia', dan 'tokopediacare'. *Query* tersebut merupakan akun resmi dari *e-commerce* Bukalapak, Lazada, dan Tokopedia. *Tweet* yang diambil merupakan postingan dari pelanggan ketiga *e-commerce* tersebut. Diagram alur proses pengambilan *dataset* seperti pada gambar 3.2.



Gambar 3.2 Diagram Proses Dataset

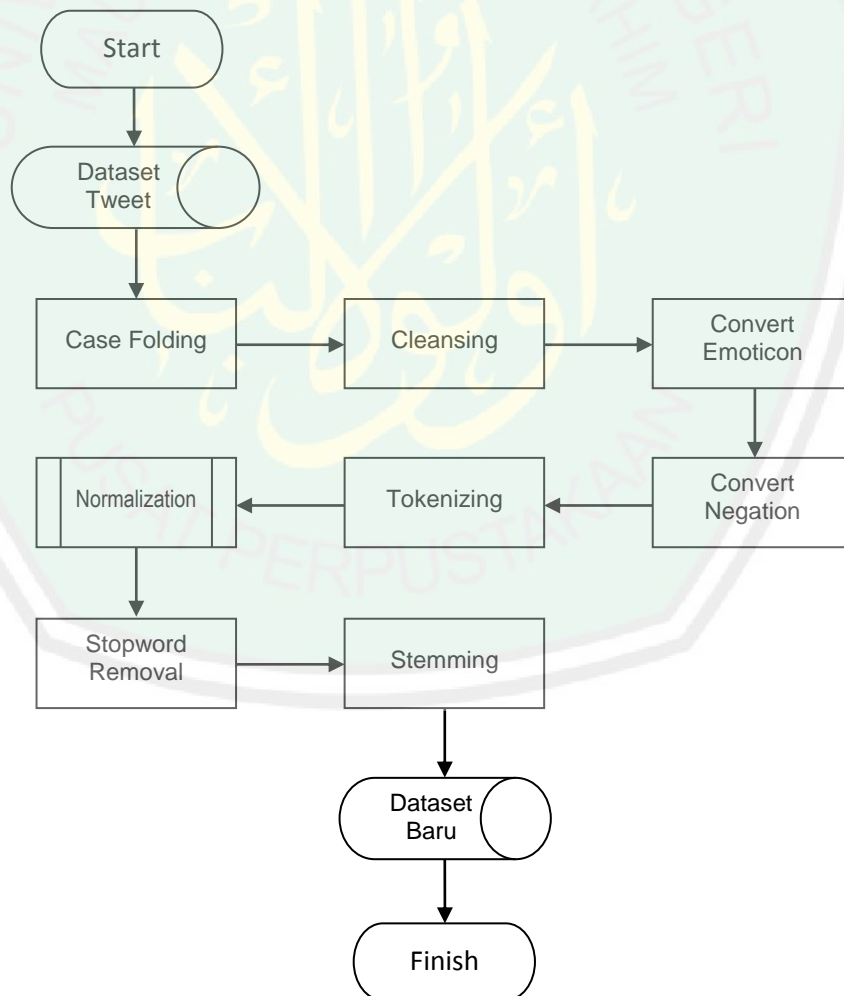
Dataset dari hasil *crawling* ini akan dibagi menjadi dua bagian yaitu data *training* dan data *testing* yang dipresentasikan pada gambar 3.2. Data hasil *crawling* ini berupa dokumen *tweet* yang tidak menyertakan atribut lainnya. Data *training* ini akan dimasukkan ke dalam *database* MySQL, dan diklasifikasikan secara manual dengan label sentimen positif atau negatif. Data *testing* yang diperoleh dari proses *crawling* dengan *APITwitter* ini akan disimpan didalam *database* MySQL, yang nantinya akan diolah kedalam sistem untuk menghasilkan *output* otomatis berupa sentimen positif atau negatif.

Pengambilan data *training* dan data *testing* dilakukan dengan cara yang berbeda, serta waktu pengambilannya berbeda. Data *training* diambil selama tiga

bulan pada bulan Januari, Februari, dan Maret dengan cara manual sedangkan data *testing* diambil secara *realtime* pada saat proses berlangsung.

3.4 Preprocessing

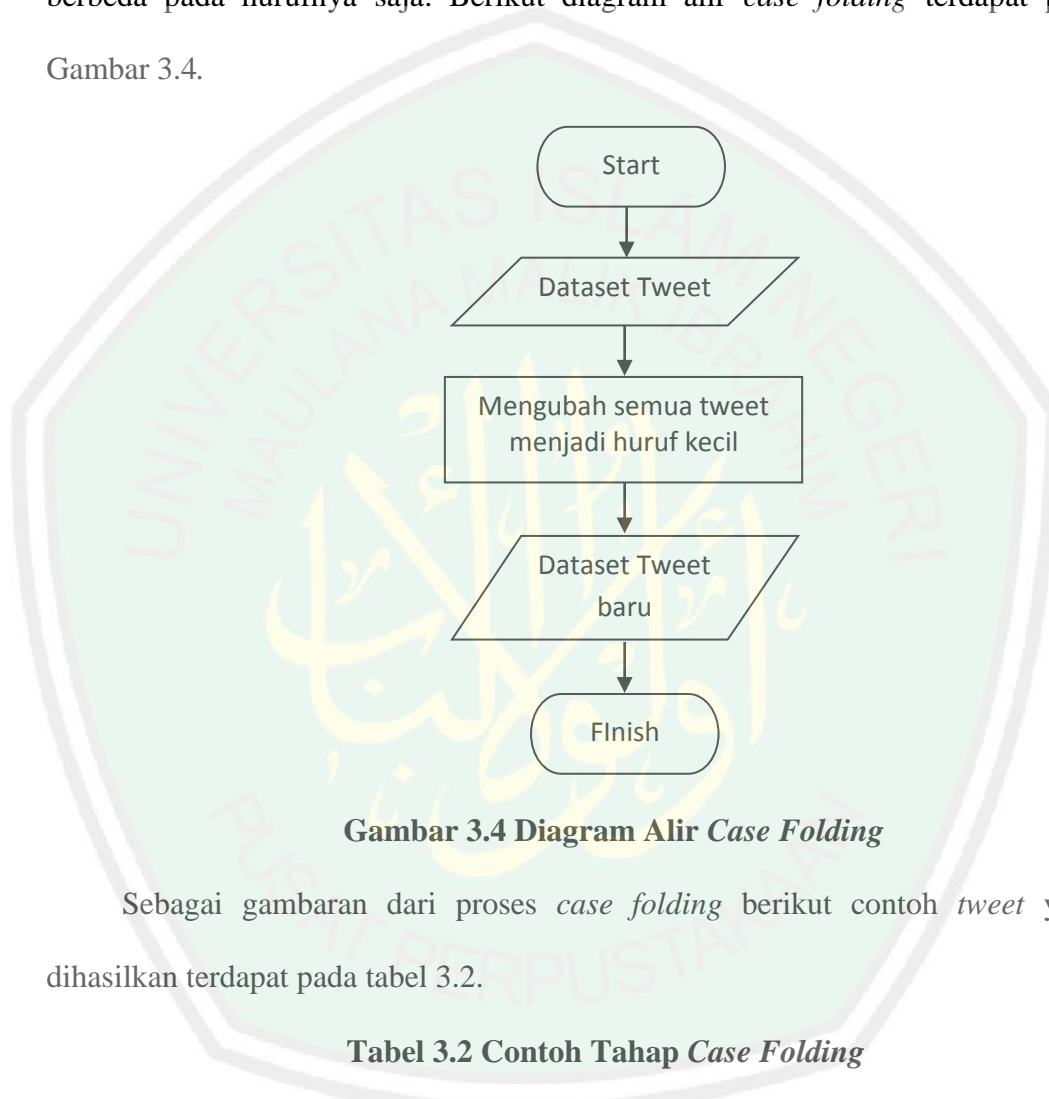
Proses *preprocessing* merupakan hal yang penting untuk tahap selanjutnya, yaitu mengurangi atribut yang kurang berpengaruh terhadap proses klasifikasi. Data yang dimasukkan pada tahap ini masih berupa data mentah yang masih kotor, sehingga hasil dari proses ini adalah dokumen yang berkualitas yang harapannya mempermudah dalam proses klasifikasi. Proses *preprocessing* terdiri dari beberapa tahapan yang dapat dilihat pada gambar 3.3 berikut.



Gambar 3.3 Diagram Alir Preprocessing

3.4.1 Case Folding

Pada tahap *case folding* huruf kapital pada semua dokumen *tweet* diubah menjadi huruf kecil. Tujuannya untuk menghilangkan *redudansi* data yang hanya berbeda pada hurufnya saja. Berikut diagram alir *case folding* terdapat pada Gambar 3.4.



Gambar 3.4 Diagram Alir Case Folding

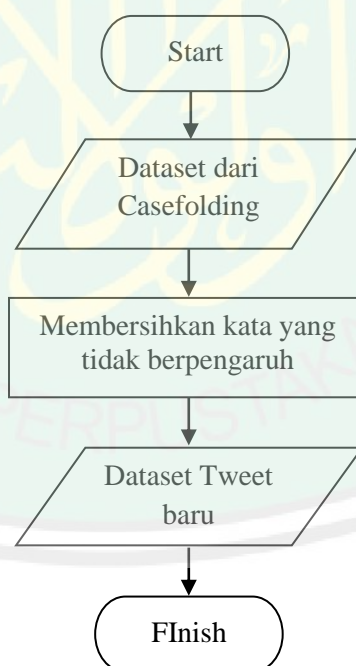
Sebagai gambaran dari proses *case folding* berikut contoh *tweet* yang dihasilkan terdapat pada tabel 3.2.

Tabel 3.2 Contoh Tahap Case Folding

Input Process	Output Process
@Yoghaaaaa Hi, pesanan kamu sudah dalam pengiriman. Maksimal kamu menerima produk 16/02/2017. Mohon ditunggu ya. Thanks –Bella	@yoghaaaaa hi, pesanan kamu sudah dalam pengiriman. maksimal kamu menerima produk 16/02/2017. mohon ditunggu ya. thanks –bella

3.4.2 *Cleansing*

Tahapan *cleansing* merupakan tahap pembersihan kata yang tidak berpengaruh sama sekali terhadap hasil klasifikasi sentimen. Komponen dokumen *tweet* memiliki berbagai atribut yang tidak berpengaruh terhadap sentimen, karena setiap *tweet* hampir semua memiliki atribut tersebut. Contoh dari atribut yang tidak penting tersebut adalah yaitu *mention* yang diawali dengan atribut ('@'), *hashtag* yang diawali dengan atribut ('#'), *link* yang diawali dengan atribut ('http', 'bit.ly') dan karakter simbol (~!@#\$\$%^&*()_+?<>.,:~?{}|[]). Atribut yang tidak berpengaruh tersebut akan dihilangkan dari dokumen kemudian akan digantikan dengan karakter spasi. Berikut diagram alir *cleansing* terdapat pada gambar 3.5 .



Gambar 3.5 Diagram Alir *Cleansing*

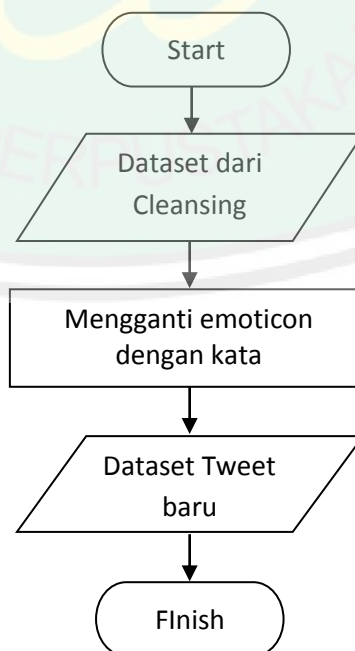
Sebagai gambaran dari proses *cleansing* berikut penulis memberikan contoh *tweet* yang dihasilkan seperti pada tabel 3.3.

Tabel 3.3 Contoh Tahap *Cleansing*

Input Process	Output Process
@yoghaaaaa hi, pesanan kamu sudah dalam pengiriman. maksimal kamu menerima produk 16/02/2017. mohon ditunggu ya. thanks –bella	hi, pesanan kamu sudah dalam pengiriman maksimal kamu menerima produk 16/02/2017 mohon ditunggu ya thanks –bella

3.4.3 *Convert Emoticon*

Tahap *convert emoticon* ini sangat berpengaruh terhadap sentimen suatu dokumen, karena *emoticon* dapat menggambarkan perasaan seseorang ketika senang atau sedih. Karena *emoticon* berupa simbol maka tahap *convert emoticon* ini akan mengubah simbol senang menjadi kata ‘*emotsenang*’ dan mengubah simbol sedih dengan kata ‘*emotsedih*’. Berikut diagram alir *convert emoticon* pada Gambar 3.6.

Gambar 3.6 Diagram Alir *Convert Emoticon*

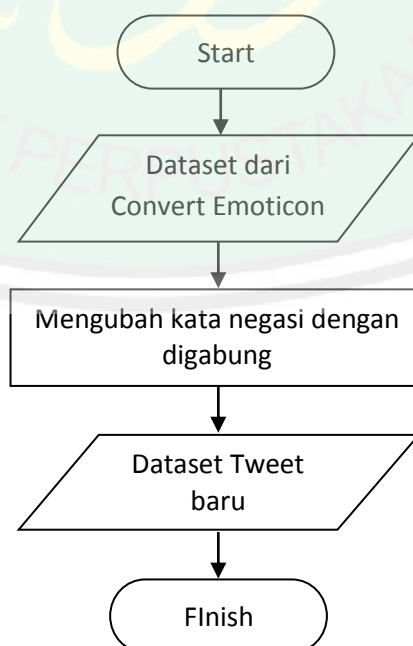
Sebagai gambaran dari proses *convert emoticon* berikut contoh *tweet* yang dihasilkan terdapat pada tabel 3.4.

Tabel 3.4 Contoh Tahap *Convert Emoticon*

Input Process	Output Process
hi, maaf ya :(bisa infokan atas nama pemesan dan nomor pesanan kamu melalui dm kami ya	hi, maaf ya emotsedih bisa infokan atas nama pemesan dan nomor pesanan kamu melalui dm kami ya

3.4.4 *Convert Negation*

Tahap *convert negation* merupakan proses konversi kata – kata negasi yang terdapat pada suatu *tweet*. Kata negasi akan merubah makna sentimen suatu dokumen, sehingga kata negasi akan digabungkan dengan kata selanjutnya. Contoh kata negasi adalah ‘bukan’, ‘tidak’, ‘jangan’ dan lain sebagainya. Berikut diagram alir *convert negation* terdapat pada gambar 3.7.



Gambar 3.7 Diagram Alir *Convert Negation*

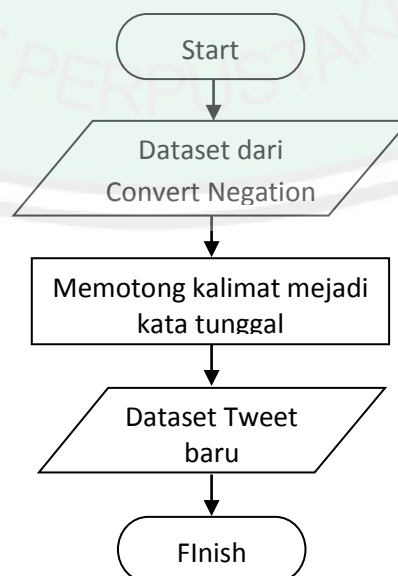
Sebagai gambaran dari proses *convert negation* berikut contoh *tweet* yang dihasilkan seperti pada tabel 3.5.

Tabel 3.5 Contoh Tahap *Convert Negation*

Input Process	Output Process
hi, mohon maaf atas	hi, mohon maaf atas
ketidaknyamanya, resi yang diberikan	ketidaknyamanya, resi yang diberikan
pelapak tidak valid, kami sudah	pelapak tidakvalid, kami sudah
melakukan invalid resi (1) ^bl	melakukan invalid resi (1) ^bl

3.4.5 *Tokenizing*

Tahap *tokenizing* merupakan pemotongan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal. Kata dalam dokumen yang dimaksud adalah kata yang dipisah oleh spasi. Sehingga hasil dari proses ini merupakan kata tunggal yang dimasukkan ke dalam *database* untuk keperluan pembobotan. Berikut diagram alir *tokenizing* terdapat pada Gambar 3.8.



Gambar 3.8 Diagram Alir *Tokenizing*

Sebagai gambaran dari proses *tokenizing* berikut contoh *tweet* yang dihasilkan seperti pada tabel 3.6.

Tabel 3.6 Contoh Tahap *Tokenizing*

Input Process	Output Process
pesanan sudah dalam pengiriman, jadi sudah tidak bisa dibatalkan	'pesanan' 'sudah' 'dalam' 'pengiriman' 'jadi' 'sudah' 'tidak' 'bisa' 'dibatalkan'

3.4.6 Normalization

Pada tahap *normalization* ini dilakukan perubahan kata yang tidak sesuai dengan EYD, sehingga dapat mengurangi hasil sentimen dokumen. Tahap ini dibagi menjadi tiga langkah, yaitu konversi kata singkatan, konversi kata baku, dan konversi kata Inggris. Berikut diagram alir *normalization* pada Gambar 3.9.



Gambar 3.9 Diagram Alir *Normalization*

Pengguna *twitter* dibatasi hanya bisa melakukan posting 140 karakter saja, sehingga menyebabkan banyak pengguna yang menulis dengan kata singkatan agar apa yang ditulisnya dapat terekspresikan. Hal tersebut membuat masalah terhadap performasi sentimen dokumen. Pada tahap ini dilakukan proses untuk mengubah kata singkatan menjadi kata normal, yang dicontohkan pada tabel 3.7.

Tabel 3.7 Contoh Tahap Konversi Kata Singkatan

Input Process	Output Process
hi, bukalapak utk sementara tdk dpt diakses & masih dalam proses utk memulihkannya mohon maaf atas ketidaknyamanannya ^bl	hi bukalapak untuk sementara tidak dapat diakses amp masih dalam proses untuk memulihkannya mohon maaf atas ketidaknyamanannya bl

Proses selanjutnya adalah tahap konversi kata baku, karena standar bahasa yang dipakai pada penelitian ini adalah bahasa Indonesia baku. Proses ini berguna untuk mengubah kata yang tidak standar bahasa Indonesia menjadi bahasa Indonesia yang baku. Terkadang pengguna *twitter* masih menggunakan bahasa daerahnya dan bahasa gaul sehingga perlu dilakukan konversi kata baku, seperti yang dicontohkan pada tabel 3.8.

Tabel 3.8 Contoh Tahap Konversi Kata Baku

Input Process	Output Process
hai yunni mohon maaf mohon informasikan terlebih dahulu email akun tokopedia via dm	hai yunni mohon maaf mohon informasi lebih dahulu email akun via

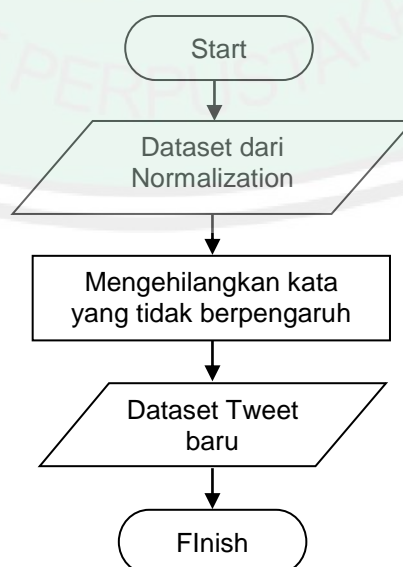
Tahap yang terakhir dalam proses *normalization* adalah mengubah kata Inggris ke Indonesia, berikut contoh konversi kata Inggris pada tabel 3.9.

Tabel 3.9 Contoh Tahap Konversi Kata Inggris

Input Process	Output Process
ulasan customer service yang menggunakan gosend hari ini thanks buat bintang nya	ulasan customer service yang menggunakan gosend hari ini terimakasih buat bintang nya

3.4.7 Stopword Removal

Tahap *stopword removal* merupakan tahap menghilangkan kata yang tidak sesuai dengan topik dokumen, jika ada kata tersebut tidak mempengaruhi akurasi dalam klasifikasi sentimen dokumen. Kata yang akan dihilangkan dihimpun dalam *database* kata *stopword*. Jika dalam dokumen *tweet* ada yang sesuai dengan kata dalam *stopword* maka kata tersebut akan dihilangkan dan diganti dengan karakter spasi. Berikut diagram alir *stopword removal* pada Gambar 3.10.



Gambar 3.10 Diagram Alir Stopword Removal

Sebagai gambaran dari proses *stopword removal* berikut contoh *tweet* yang telah dilakukan proses *stopword removal* akan menghasilkan dokumen seperti pada tabel 3.10.

Tabel 3.10 Contoh Tahap *Stopword Removal*

Input Process	Output Process
bukan kah resinya sudah saya kirim dan juga sudah di konfirm oleh lazada klo barangnya sudah sampai di gudang lazada	bukan resinya kirim konfirm lazada klo barangnya gudang lazada

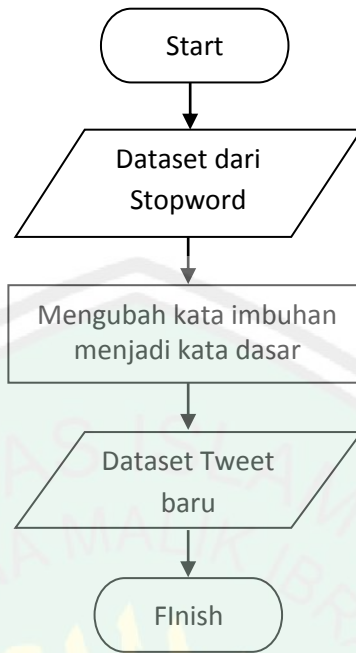
3.4.8 *Stemming*

Pada tahap *stemming* merupakan suatu proses untuk mengubah kata – kata yang terdapat dalam suatu dokumen ke dalam kata – kata akarnya dengan menggunakan aturan – aturan tertentu. Proses *stemming* bahasa Indonesia dilakukan dengan menghilangkan *sufiks*, *prefix*, dan *konfiks* pada dokumen. Pada proses *stemming* ini penulis menggunakan metode yang dibuat oleh Bobby Nazief dan Mirna Adriani, dengan tahapan sebagai berikut: (Agusta, 2009)

- 1) Cari kata yang akan *distem* dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tersebut adalah *root word*. Maka metode berhenti.
- 2) *Inflection Suffixes* (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa *particles* (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus *Possesive Pronouns* (“-ku”, “-mu”, atau “-nya”), jika ada.
- 3) Hapus *Derivation Suffixes* (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka metode berhenti. Jika tidak maka ke langkah 3a

- a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka metode berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
- 4) Hapus *Derivation Prefix*. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
- a. Periksa tabel kombinasi awalan – akhiranyang tidak diijinkan. Jika ditemukan maka metode berhenti, jika tidak pergi ke langkah 4b.
 - b. *For i = 1 to 3*, tentukan tipe awalan kemudian hapus awalan. Jika *root word* belum juga ditemukan lakukan langkah 5, jika sudah maka metode berhenti. Catatan: jika awalan kedua sama dengan awalan pertama metode berhenti.
- 5) Melakukan *recording*.
- 6) Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *root word*. Proses selesai.

Untuk membuat proses setemming secara sempurna harus melalui beberapa tahapan poroses yang telah disebutkan pada poin diatas.Berikut diagram alir *stemming* pada Gambar 3.11.



Gambar 3.11 Diagram Alir Stemming

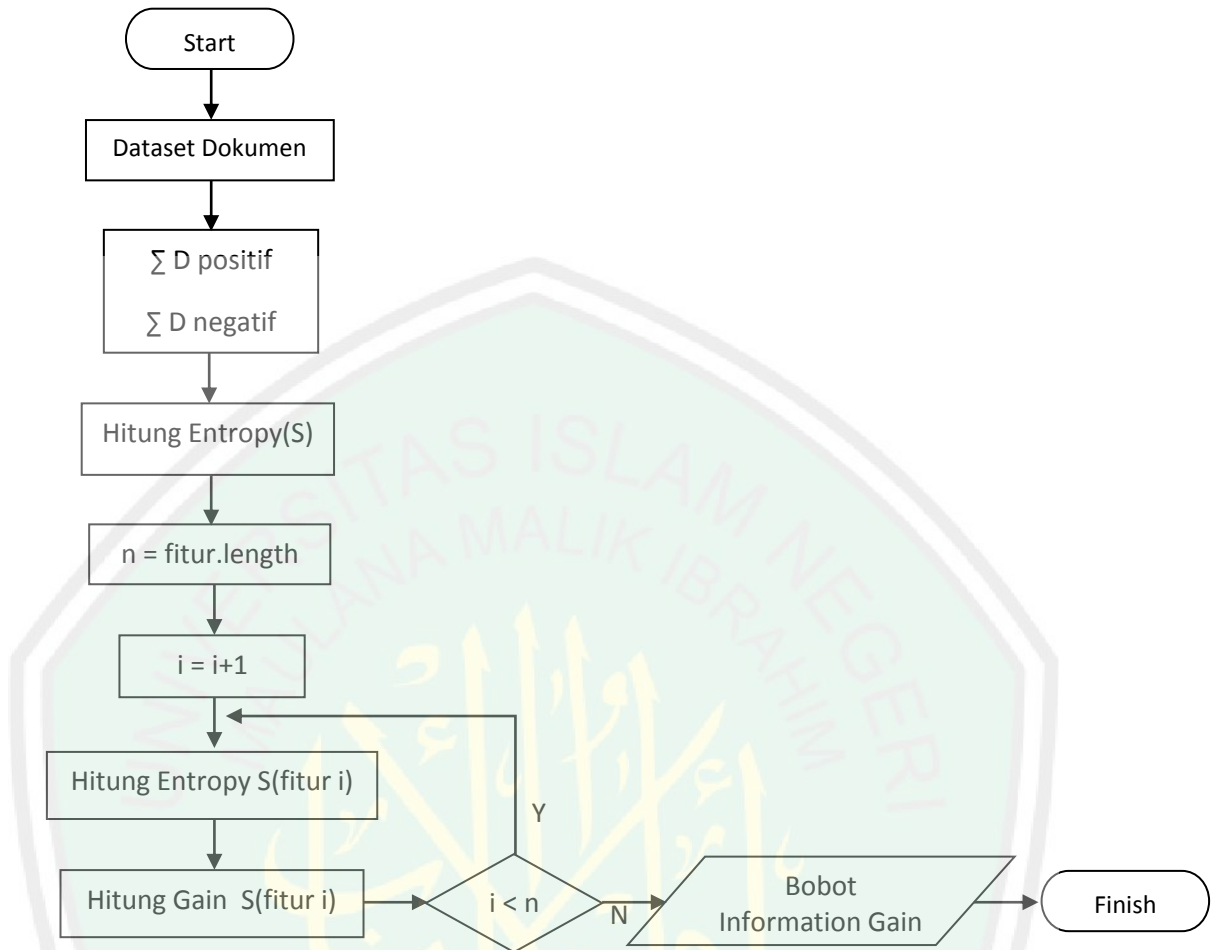
Sebagai gambaran dari proses *stemming* berikut contoh *tweet* yang dihasilkan seperti pada tabel 3.11.

Tabel 3.11 Contoh Tahap Stemming

Input Process	Output Process
kalau tetap tidak terkirim pakatnya dianggap batal gimana kalau batal pesanan min pembayaran via cod	kalau tetap tidak kirim paket anggap batal gimana pesan kalau batal pesan min bayar via cod

3.4.9 Seleksi *Fitur Information Gain*

Pada proses seleksi *fitur* ini merupakan proses untuk penyeleksian *fitur* yang paling penting dalam proses klasifikasi agar akurasi dari sistem bisa lebih baik. Berikut *flowchart Information Gain* pada gambar 3.12.



Gambar 3.12 Diagram Alir Metode *Information Gain*

Pada pembahasan berikut penulis memberikan contoh perhitungan *information gain* secara manual. Dengan cara menghitung bobot sesuai rumus 2.3 Pada bab 2, metode *information gain* secara sederhana dapat dicontohkan seperti tabel 3.12.

Tabel 3.12 Contoh Koleksi Data

Dokumen	Fitur			Sentimen
	Puas	Terimakasih	Batal	
D ₁	Ya	Ya	Tidak	P
D ₂	Ya	Ya	Tidak	P
D ₃	Ya	Ya	Tidak	N

D ₄	Tidak	Ya	Ya	P
D ₅	Tidak	Tidak	Ya	N
D ₆	Tidak	Tidak	Ya	N
D ₇	Tidak	Tidak	Tidak	P
D ₈	Ya	Tidak	Tidak	N
D ₉	Ya	Ya	Ya	P
D ₁₀	Ya	Ya	Tidak	P

Fitur yang terdapat pada tabel 3.12 merupakan potongan kata dari dokumen yang akan dihitung bobotnya. Pada kasus perhitungan bobot *information gain* penulis mengambil contoh kata ‘puas’ dengan menghitung *entropy* pada *dataset* dengan menggunakan persamaan 2.1, sebagai berikut:

$$Entropy (Set) = - \left[\left(\frac{6}{10} \right) \log_2 \left(\frac{6}{10} \right) + \left(\frac{4}{10} \right) \log_2 \left(\frac{4}{10} \right) \right] = 0,971$$

Selanjutnya ambil contoh pada kata ‘puas’ yang memiliki *value* Ya atau Tidak sehingga dapat dihitung dengan persamaan 2.1 setelah itu hitung *Entropy* (S_{puas}) dengan persamaan 2.2 dan menghasilkan perhitungan sebagai berikut :

$$Entropy (Positif) = - \left[\left(\frac{4}{6} \right) \log_2 \left(\frac{4}{6} \right) + \left(\frac{2}{6} \right) \log_2 \left(\frac{2}{6} \right) \right] = 0,9183$$

$$Entropy (Negatif) = - \left[\left(\frac{2}{4} \right) \log_2 \left(\frac{2}{4} \right) + \left(\frac{2}{4} \right) \log_2 \left(\frac{2}{4} \right) \right] = 1$$

$$Entropy (S_{puas}) = \left(\frac{6}{10} \right) 0,9183 + \left(\frac{4}{10} \right) \times 1 = 0,95098$$

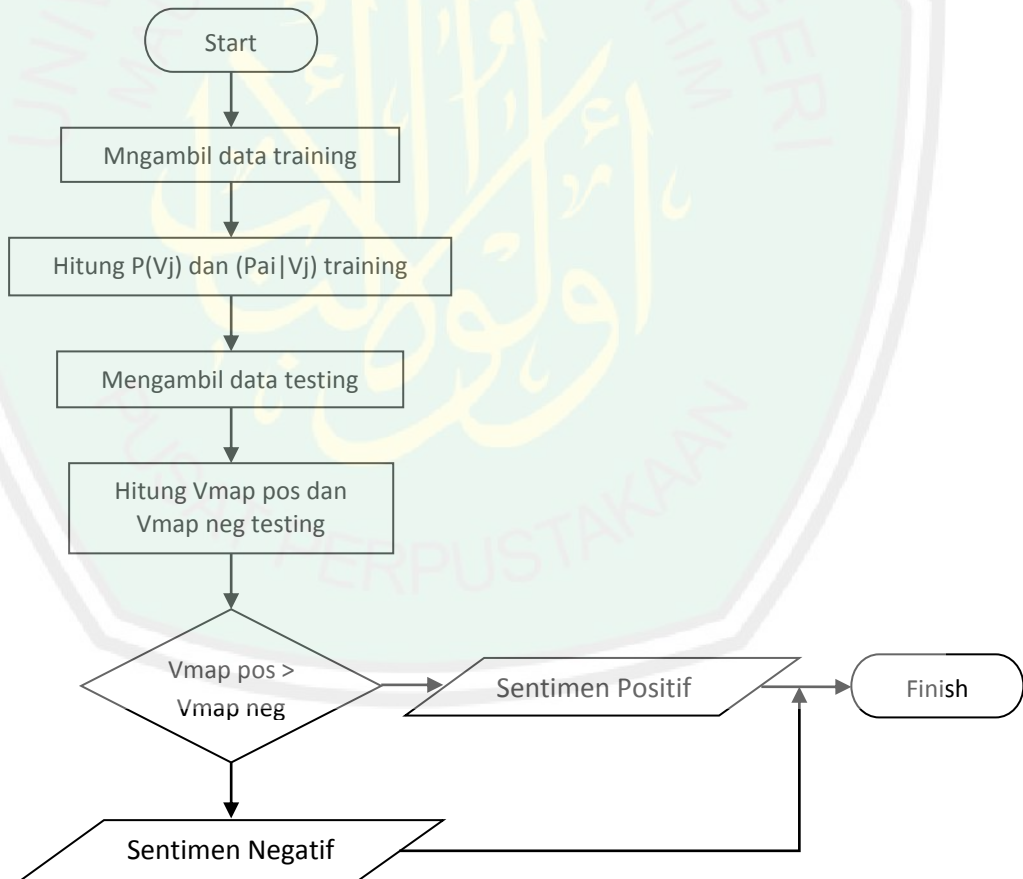
Langkah terakhir untuk mencari nilai *information gain* menggunakan persamaan 2.3 sebagai berikut :

$$Gain(S_{puas}) = 0,971 - 0,95098 = 0,02002$$

Dengan bobot *information gain* tersebut setiap kata akan dirangking dan hasilnya akan didapatkan *fitur* yang terbaik. Untuk keperluan klasifikasi data *testing* nantinya akan diambil fitur dengan akurasi tertinggi.

3.5 Naïve Bayes Classification

Pada tahap klasifikasi ini menggunakan metode *Naïve Bayes* yang dibagi menjadi dua proses, yaitu proses *training* dan proses *testing*. Pada tahap ini dilakukan proses *training* terlebih dahulu untuk pelatihan, selanjutnya dilakukan proses *testing* dengan mengacu *probabilitas* dari *dataset training*. Diagram alur tahapan klasifikasi *Naïve Bayes* ini ditunjukkan pada gambar 3.13 berikut.



Gambar 3.13 Diagram Alir Metode *Naïve Bayes*

Pada bahasan berikut penulis membuat contoh perhitungan manual *Naïve Bayes Classification* dengan sampel 6 *tweet* data taining dan 3 *tweet* data testing.

3.5.1 Proses *Training*

Sebuah dokumen *training* sudah diklasifikasikan secara manual dan sudah dilakukan proses *preprocessing* seperti pada tabel 3.13 berikut:

Tabel 3.13 Contoh Kasus Data *Training*

Tweet	Fitur	Kategori
Tweet1	sudah kami bantu remit silakan cek lebih lanjut	P
Tweet2	baik, tunggu. terimakasih banyak	P
Tweet3	silakan kami sudah respon pesan terimakasih	P
Tweet4	mohon maaf, resi tidak valid	N
Tweet5	tidak kenal, email tidak ada	N
Tweet6	mohon maaf, mohon informasi	N

Dari data tabel 3.13 dibuat sebuah model *probabilitas* dengan mengacu pada persamaan 2.10 dan 2.11 sebagai berikut:

$$P(a_{\text{terimakasih}} | V_{\text{positif}}) = \frac{2 + 1}{18 + 23} = \frac{3}{41}$$

$$P(a_{\text{terimakasih}} | V_{\text{negatif}}) = \frac{0 + 1}{14 + 23} = \frac{1}{37}$$

Jika dibuat menjadi sebuah tabel, maka *probabilitas* setiap kata pada data *training* seperti pada tabel 3.14 berikut:

Tabel 3.14 Perhitungan Probabilitas Data *Training*

Kategori	P(v _j)	P(a _i V _j)							
		sudah	kami	Bantu	remit	Silakan	Cek	lebih	lanjut
P	$\frac{1}{2}$	$\frac{3}{41}$	$\frac{3}{41}$	$\frac{2}{41}$	$\frac{2}{41}$	$\frac{3}{41}$	$\frac{2}{41}$	$\frac{2}{41}$	$\frac{2}{41}$
N	$\frac{1}{2}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$

Kategori	P(v _j)	P(a _i V _j)						
		Baik	tunggu	terimakasih	Banyak	respon	pesan	mohon
P	$\frac{1}{2}$	$\frac{2}{41}$	$\frac{2}{41}$	$\frac{3}{41}$	$\frac{2}{41}$	$\frac{2}{41}$	$\frac{2}{41}$	$\frac{1}{41}$
N	$\frac{1}{2}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{1}{37}$	$\frac{4}{37}$

Kategori	P(v _j)	P(a _i V _j)							
		maaf	resi	tidak	Valid	kenal	Email	ada	informasi
P	$\frac{1}{2}$	$\frac{1}{41}$	$\frac{1}{41}$	$\frac{1}{41}$	$\frac{1}{41}$	$\frac{1}{41}$	$\frac{1}{41}$	$\frac{1}{41}$	$\frac{1}{41}$
N	$\frac{1}{2}$	$\frac{3}{37}$	$\frac{2}{37}$	$\frac{4}{37}$	$\frac{2}{37}$	$\frac{2}{37}$	$\frac{2}{37}$	$\frac{2}{37}$	$\frac{2}{37}$

Hasil perhitungan *probabilitas* tersebut digunakan sebagai model probabilistik yang selanjutnya digunakan sebagai data acuan untuk menentukan data *testing*.

3.5.2 Proses *Testing*

Berikut penulis beri contoh perhitungan untuk data *testing*, dengan data *testing* seperti pada tabel 3.15.

Tabel 3.15 Contoh Kasus Data Testing

Tweet	<i>Fitur</i>	Kategori
Tweet7	terimakasih telah pesan barang lapak	?
Tweet8	silakan tunggu hadiah telah pesan	?
Tweet9	maaf tidak bisa kirim besok	?

Proses *testing* ini dihitung *probabilitasnya* dan dicari *probabilitas* tertinggi menggunakan persamaan 2.9 sebagai berikut:

$$\begin{aligned}
& P(\text{Tweet7}|V_{\text{positif}}) \\
&= P(a_{\text{terimakasih}}|V_{\text{positif}}) \times P(a_{\text{telah}}|V_{\text{positif}}) \times P(a_{\text{pesan}}|V_{\text{positif}}) \times \\
&\quad P(a_{\text{barang}}|V_{\text{positif}}) \times P(a_{\text{lapak}}|V_{\text{positif}}) \times P(V_{\text{positif}}) \\
&= \frac{3}{41} \times 1 \times \frac{2}{41} \times 1 \times 1 \times \frac{1}{2} \\
&= 0,001785
\end{aligned}$$

$$\begin{aligned}
& P(\text{Tweet7}|V_{\text{negatif}}) \\
&= P(a_{\text{terimakasih}}|V_{\text{negatif}}) \times P(a_{\text{telah}}|V_{\text{negatif}}) \times P(a_{\text{pesan}}|V_{\text{negatif}}) \times \\
&\quad P(a_{\text{barang}}|V_{\text{negatif}}) \times P(a_{\text{lapak}}|V_{\text{negatif}}) \times P(V_{\text{negatif}}) \\
&= \frac{1}{37} \times 1 \times \frac{1}{37} \times 1 \times 1 \times \frac{1}{2} \\
&= 0,00036523
\end{aligned}$$

$$\begin{aligned}
& P(\text{Tweet8}|V_{\text{positif}}) \\
&= P(a_{\text{silakan}}|V_{\text{positif}}) \times P(a_{\text{tunggu}}|V_{\text{positif}}) \times P(a_{\text{hadiah}}|V_{\text{positif}}) \times \\
&\quad P(a_{\text{telah}}|V_{\text{positif}}) \times P(a_{\text{pesan}}|V_{\text{positif}}) \times P(V_{\text{positif}}) \\
&= \frac{3}{41} \times \frac{2}{41} \times 1 \times 1 \times \frac{2}{41} \times \frac{1}{2} \\
&= 0,000087056
\end{aligned}$$

$$\begin{aligned}
& P(\text{Tweet8}|V_{\text{negatif}}) \\
&= P(a_{\text{silakan}}|V_{\text{negatif}}) \times P(a_{\text{tunggu}}|V_{\text{negatif}}) \times P(a_{\text{hadiah}}|V_{\text{negatif}}) \times \\
&\quad P(a_{\text{telah}}|V_{\text{negatif}}) \times P(a_{\text{pesan}}|V_{\text{negatif}}) \times P(V_{\text{negatif}}) \\
&= \frac{1}{37} \times \frac{1}{37} \times 1 \times 1 \times \frac{1}{37} \times \frac{1}{2} \\
&= 0,000009871
\end{aligned}$$

$$\begin{aligned}
& P(\text{Tweet9} | V_{\text{positif}}) \\
&= P(a_{\text{maaf}} | V_{\text{positif}}) \times P(a_{\text{tidak}} | V_{\text{positif}}) \times P(a_{\text{bisa}} | V_{\text{positif}}) \times \\
&\quad P(a_{\text{kirim}} | V_{\text{positif}}) \times P(a_{\text{besok}} | V_{\text{positif}}) \times P(V_{\text{positif}}) \\
&= \frac{1}{41} \times \frac{1}{41} \times 1 \times 1 \times 1 \times \frac{1}{2} \\
&= 0,000297442
\end{aligned}$$

$$\begin{aligned}
& P(\text{Tweet9} | V_{\text{negatif}}) \\
&= P(a_{\text{maaf}} | V_{\text{negatif}}) \times P(a_{\text{tidak}} | V_{\text{negatif}}) \times P(a_{\text{bisa}} | V_{\text{negatif}}) \times \\
&\quad P(a_{\text{kirim}} | V_{\text{negatif}}) \times P(a_{\text{besok}} | V_{\text{negatif}}) \times P(V_{\text{negatif}}) \\
&= \frac{3}{37} \times \frac{2}{37} \times 1 \times 1 \times 1 \times \frac{1}{2} \\
&= 0,00219138
\end{aligned}$$

Setelah menghitung *probabilitas* dari setiap data tester, diperoleh hasil pada tabel 3.16 sebagai berikut:

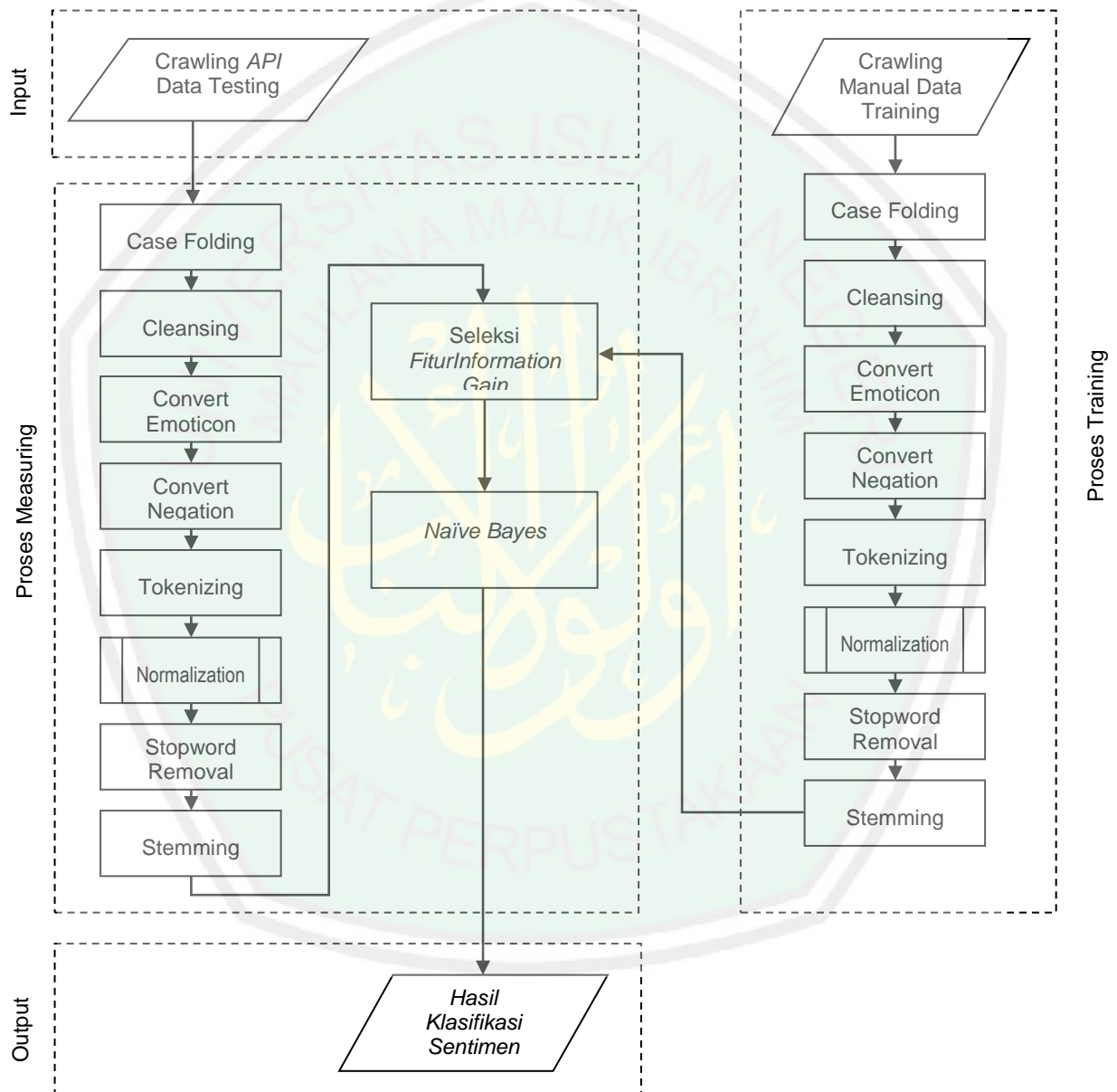
Tabel 3.16 Nilai Probabilitas Data Testing

Tweet	Probabilitas	
	Positif	Negatif
Tweet7	0,001785	0,00036523
Tweet8	0,000087056	0,000009871
Tweet9	0,000297442	0,00219138

Pada tabel 3.16 dapat dianalisis hasil dari data *testing* pertama yaitu *Tweet7* didapatkan nilai *probabilitas* positif lebih besar dari nilai *probabilitas* negatif sehingga dapat disimpulkan bahwa *Tweet7* termasuk kategori sentimen positif. Pada *Tweet8* didapatkan nilai *probabilitas* positif lebih besar dari nilai

probabilitas negatif sehingga *Tweet8* termasuk kategori sentimen positif. Sedangkan pada *Tweet9* nilai *probabilitas* negatif lebih besar dari nilai *probabilitas* positif sehingga *Tweet9* termasuk kategori sentimen negatif.

3.6 Perancangan Desain Sistem

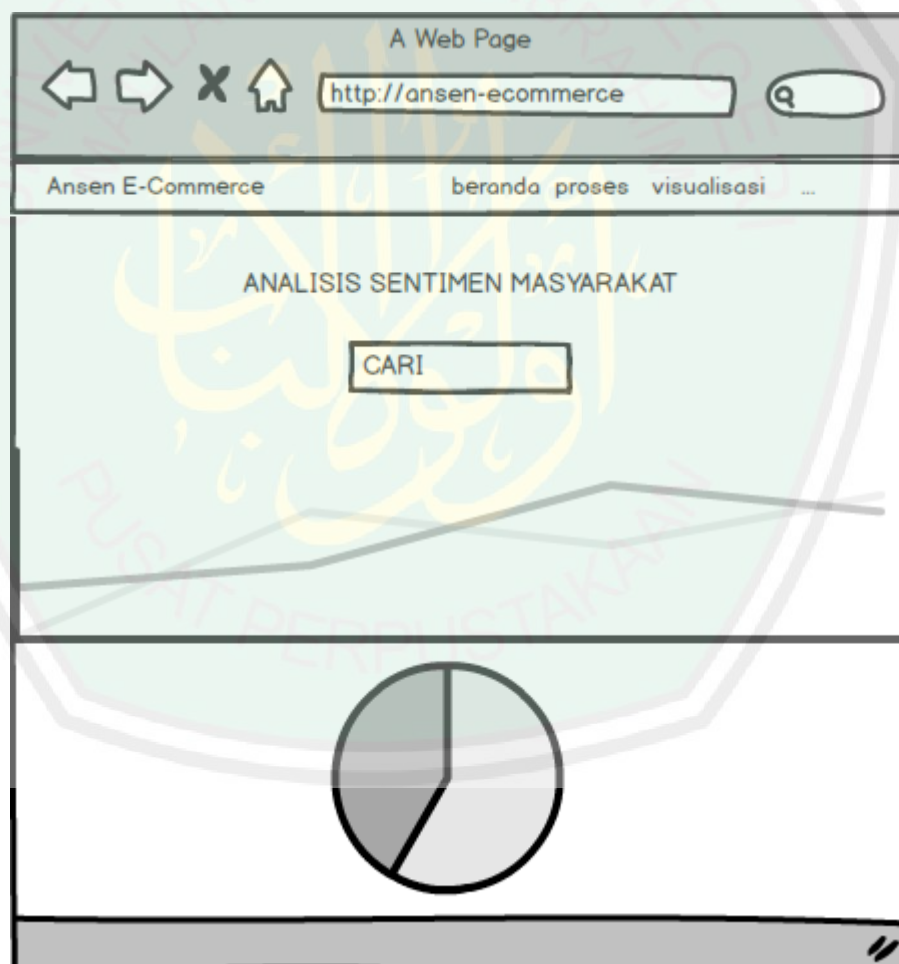


Gambar 3.14 Desain Sistem

3.7 Perancangan Antar Muka

Pada subbab ini membahas mengenai perancangan antar muka atau *interface system* yang dibangun. Terdapat antar muka yang memiliki fungsional masing – masing. Antar muka tersebut adalah halaman utama, proses *crawling*, proses *preprocessing*, proses bobot *Information Gain*, proses bobot *Naïve Bayes*, proses klasifikasi *Naïve Bayes*, dan terakhir visualisasi. Berikut tampilan dari masing – masing yang telah disebutkan.

a) Antar Muka Halaman Utama



Gambar 3.15 Tampilan Halaman Utama

b) Antar Muka Proses *Crawling*

Id tweet	Tweet	Tanggal
184937	ini adalah tweet dari twitter	12-09-2017
184937	ini adalah tweet dari twitter	12-09-2017
184937	ini adalah tweet dari twitter	12-09-2017

Gambar 3.16 Tampilan *Crawling*c) Antar Muka Proses *Preprocessing*

Id twe	Tweet	Tweet Preprocess	Tanggal
18493	ini adalah tweet dari	Hasil preprocessing	12-09-2
18493	ini adalah tweet dari	Hasil preprocessing	12-09-2
18493	ini adalah tweet dari	Hasil preprocessing	12-09-2

Gambar 3.17 Tampilan *Preprocessing*

d) Antar Muka Proses Bobot *Information Gain*

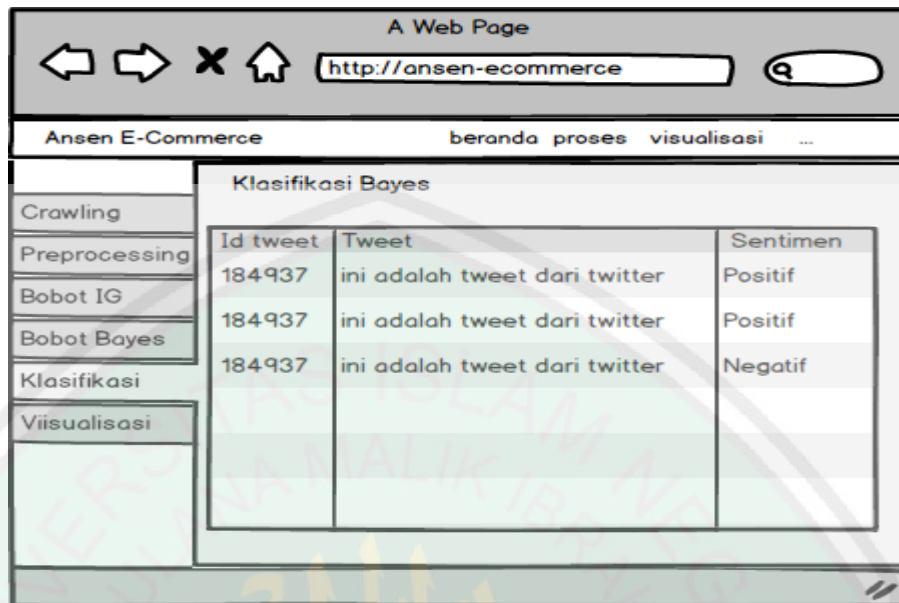
Id kata	Kata	Entropy Kata	Information Gain
1	iya	0.8456245	0.2935279
2	bagus	0.8456245	0.2935279
3	ajib	0.8456245	0.2935279
4	keren	0.8456245	0.2935279

Gambar 3.18 Tampilan Bobot *Information Gain*

e) Antar Muka Proses Bobot *Naïve Bayes*

Id twee	Tweet	Prob Positif	Prob Negatif
184937	ini adalah tweet dari tw	0.2352352	0.2435252
184937	ini adalah tweet dari tw	0.2352352	0.2435252
184937	ini adalah tweet dari tw	0.2352352	0.2435252

Gambar 3.19 Tampilan Bobot *Naïve Bayes*

f) Antar Muka Proses Klasifikasi *Naïve Bayes*


Id tweet	Tweet	Sentimen
184937	ini adalah tweet dari twitter	Positif
184937	ini adalah tweet dari twitter	Positif
184937	ini adalah tweet dari twitter	Negatif

Gambar 3.20 Tampilan Klasifikasi *Naïve Bayes*

g) Antar Muka Visualisasi



Gambar 3.21 Tampilan Visualisasi

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil dan pembahasan sistem yang telah dibangun. Untuk itu, dilakukan pengujian terhadap sistem, dengan cara membuat proses *crawling*, proses *preprocessing*, proses klasifikasi dengan *Naïve Bayes*, proses akurasi sistem yang dibangun, dan membuat suatu visualisasi data.

4.1 Implementasi

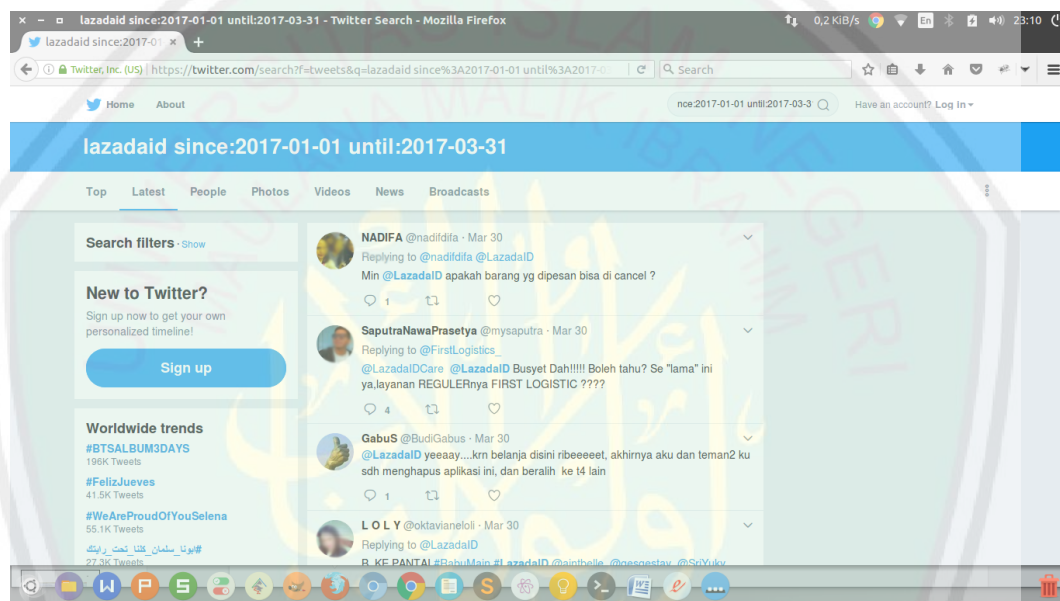
Peneilitian Analisis Sentimen ini diimplementasikan pada platform web browser, yaitu menggunakan bahasa pemrograman PHP dan didesain dengan HTML, CSS, dan Javascript, serta menggunakan MySQL sebagai *Database Management System*. Aplikasi ini dijalankan menggunakan Sistem Operasi Linux Ubuntu 64-bit yang mempunyai spesifikasi *processor* Intel® Core™ i5-3230M CPU @ 2.60GHz × 4 dan *memory* RAM sebesar 7,7 GiB. Untuk koneksi internet menggunakan jaringan 3,5G berkecepatan hingga 7,2 Mbps.

Implementasi metode ini dimulai dengan tahap pengumpulan data, melakukan sentimen manual, menghitung bobot kata hingga mendapatkan klasifikasi sentimen secara otomatis. Beberapa tahapan dalam proses klasifikasi pada penelitian ini dapat diuraikan sebagai berikut.

4.1.1 Pengumpulan Data *Training*

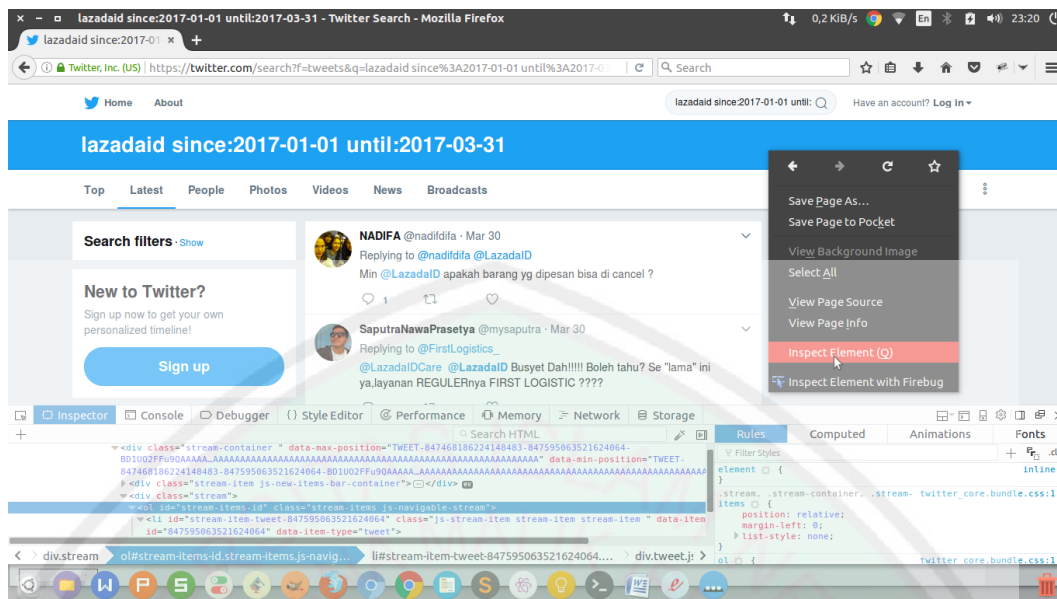
Pengumpulan data dibedakan menjadi dua tahap, pada tahap pertama ini disebut proses *training*. Data yang diambil masih menggunakan cara manual menggunakan *fitur inspect element* pada web browser. Langkah awal adalah mengunjungi situs pencarian Twitter selanjutnya menentukan kata kunci akun

tweet yang akan di *crawling* dan menentukan rentang tanggal, tulis pada kolom pencarian dengan kata kunci yang diinginkan. Pada penelitian kali ini penulis menggunakan kata kunci ‘lazadaid’, ‘lazadaid_care’, ‘bukalapak’, ‘bukalapak_care’, ‘tokopedia’, dan ‘tokopedia_care’ dengan membatasi rentang tanggal posting antara 1 Januari 2017 sampai 31 Maret 2017. Untuk lebih jelasnya bisa dilihat pada gambar 4.1 berikut.



Gambar 4.1 Pencarian Kata Kunci

Langkah selanjutnya adalah melakukan *scrolling* halaman web ke bawah sampai berhenti di tanggal yang telah ditentukan, gunanya untuk memuat semua *tweet* yang telah ditentukan. Setelah semua *tweet* muncul pada halaman web browser, klik kanan pada halaman web dan pilih *inspect element* yang gunanya adalah melihat semua halaman web lengkap beserta tag – tag elementnya. Pada menu ini didapatkan fitur untuk menyalin semua halaman tag html beserta isinya. Berikut lebih jelasnya dapat dilihat pada gambar 4.2.



Gambar 4.2 FiturInspect Element

Setelah tampil menu *inspect element* pada tab *Inspector* cari tag ‘`<ol id="stream-items-id".....`’ lalu klik kanan dan pilih *copyouter HTML*. Kemudian *paste* pada notepad dan simpan dengan extentsi *.txt*. Selanjutnya data txt tersebut dibersihkan tag – tag HTML-nya sehingga didapat *tweet* dan tanggal postingnya. Langkah berikutnya export file txt tersebut ke dalam *database MySQL*.

Untuk data training, dipilih sebanyak 3000 *tweet* secara acak selama 3 bulan yaitu bulan Januari sampai Maret 2017. Dengan memberikan sentimen positif atau negatif secara manual pada setiap *tweet*-nya. Rincian data *training* dapat dilihat pada tabel 4.12 berikut.

Tabel 4.1 Rincian Jumlah Data *Training*

Kata Kunci	Sentimen Positif	Sentimen Negatif	Total Sentimen
Lazada	342	562	904
Bukalapak	339	609	948
Tokopedia	348	800	1148

4.1.2 Pengumpulan Data *Testing*

Pengumpulan data *testing* berbeda dengan data *training*, pada proses ini data diambil dari situs jejaring sosial *Twitter* dengan cara *crawling* menggunakan API *twitter*. *Crawling* data *testing* dilakukan secara *realtime* sehingga data yang diambil adalah data yang terbaru, dan proses ini membutuhkan koneksi internet untuk mengambil data dari server *twitter*.

Dalam pengambilan data *testing* langkah pertama harus masuk akun *twitter* untuk mendapatkan token. Cara mendapatkan token masuk pada website <https://apps.twitter.com/app/new> lalu mengisi kolom pendaftaran sesuai ketentuan oleh pihak *twitter*. Token yang telah didapat dari *twitter* ini dimasukkan kedalam fungsi *get_tweet* dalam bahasa program PHP untuk koneksi API *twitter* sehingga user mendapatkan hak akses secara *realtime* dalam pencarian *tweet*. Gambaran fungsi *get_tweet* yang berisi kode token pada program PHP adalah sebagaimana pada gambar 4.3.

```
public function get_tweet($keyword, $id_akun) {
    $key = '2IVubz2TIA1wc2iLnUn9t7Ta9';
    $secret_key = '1IJ4v6eWRnECQjztEnKGqw7HiZnCVRceeiFX6m067ptcP7gRsg';
    $token = '432207100-C010q86QA71zBhVmC40QwRIKxnhcD3smwrWnREeh';
    $secret_token = 'mxBhaUjEFx1RIDrOa8wBDnCFoHE33HQCuLSLbgtneby3r';
    $conn = new TwitterOAuth($key, $secret_key, $token, $secret_token);
}
```

Gambar 4.3 Fungsi Koneksi dengan Token

Pada tahap selanjutnya jika koneksi API *twitter* berhasil maka proses *crawling* bisa dilanjutkan. Langkah awal untuk proses *crawling* data *testing* adalah menetikkan kata kunci *tweet* yang akan diproses pada kolom pencarian. Hasil kata kunci pencarian ini akan diproses pada fungsi *get_tweet* yang menghasilkan data berupa *tweet id_tweet*, *text*, dan tanggal *tweet*. *Tweet* yang didapatkan pada

saat proses *crawling* adalah *tweet* selama 7 hari kebelakang pada saat proses *crawling* dilaksanakan dan setiap harinya oleh pihak *twitter* dibatasi 100 *tweet* perhari pada setiap kata kunci. Fungsi untuk proses *crawling* dapat dilihat pada gambar 4.4.

```

for ($j=7; $j >= 0 ; $j--) {
    $tanggal_crawling = date('Y-m-d', strtotime('-'.$j.' days', strtotime(
    date('Y-m-d') ));
    $tweets = $conn->get('search/tweets', array('q'=>$keyword,
    'count'=>100, 'lang'=>'in', 'until'=>$tanggal_crawling));
    foreach ($tweets->statuses as $tweet) {
        static $i=0;
        $i+=1;
        $id_tweet = $tweet->id;
        $user = $tweet->user->screen_name;
        $text = $tweet->text;
        $date = date('Y-m-d', strtotime($tweet->created_at));
        $query = $this->con->query("INSERT INTO `data_testing` (`id_akun`,
        `id_tweet`, `tweet`, `tanggal`) VALUES (". $id_akun.", ". $id_tweet.",
        " ". $text.", " ". $date."");
    }
}

```

Gambar 4.4 Fungsi Crawling Data Testing

Ketika fungsi fungsi *get_tweet* dieksekusi maka data akan otomatis tersimpan kedalam *database* dan data siap untuk diolah pada proses selanjutnya yaitu proses *preprocessing*.

4.1.3 Proses *Preprocessing* Dokumen

Data *crawling* yang telah disimpan didalam *database* selanjutnya dilakukan proses *preprocessing* untuk menghilangkan kata yang kurang berpengaruh dalam proses klasifikasi nantinya. Proses ini membaca setiap kata yang ada dalam *database* data *training* dan data *testing* sehingga ketika selesai proses ini,

hasil dari dokumen *preprocessing* akan disimpan dalam *database*. Berikut penjelasan detail sumber kode dari setiap proses *preprocessing*.

a) *Case Folding*

Proses *case folding* merupakan proses pertama yang dilakukan dalam *preprocessing*, masukan pada proses ini adalah *tweet* yang didapat dari *database* hasil dari proses *crawling*. Fungsi *case_folding* ini akan mengubah huruf kapital menjadi huruf kecil, fungsi *case_folding* yang dibuat dapat dilihat pada gambar 4.5

```
public function case_folding($tweet){
    return strtolower($tweet);
}
```

Gambar 4.5 Fungsi Proses Case Folding

b) *Cleansing*

Proses *cleansing* dalam aplikasi berguna untuk membersihkan data tweet seperti angka, tanda baca, link, hastag, mention, dan menghasilkan kata yang akan diolah pada proses selanjutnya. Hasil proses *cleansing* adalah menghilangkan katadari karakter yang tidak ada artinya dan akan diganti oleh karakter spasi. Fungsi yang dibuat dapat dilihat pada gambar 4.6.

```
public function cleansing($tweet){
    $tweet = explode(' ', $tweet);
    $tweet_hasil = [];
    foreach ($tweet as $tweet_kata) {
        if ($tweet = preg_match('/pic.twitter.com/', $tweet_kata)) {
            $tweet_kata = "";
        } else {
            array_push($tweet_hasil, $tweet_kata);
        }
    }
}
```

```

if ($tweet = preg_match('/[0-9]/', $tweet_kata)) {
    $tweet_kata = "";

} else {
    array_push($tweet_hasil, $tweet_kata);
}

}

$tweet = implode(' ', $tweet_hasil);
$tweet = str_replace("?", " ", $tweet);
$tweet = str_replace(".", " ", $tweet);
$tweet = str_replace(",", " ", $tweet);
$tweet = str_replace(":", " ", $tweet);
$tweet = preg_replace('/@[A-Z0-9+&@#\/%?~_!$!;,.;]*[A-Z0-9+&@#\/%?~_!$!;,.;]/i', ' ', $tweet);
$tweet = preg_replace('/#[A-Z0-9+&@#\/%?~_!$!;,.;]*[A-Z0-9+&@#\/%?~_!$!;,.;]/i', ' ', $tweet);
$tweet = preg_replace('/\b(https?|ftp|file|http):\\\/[A-Z0-9+&@#\/%?~_!$!;,.;]*[A-Z0-9+&@#\/%?~_!$!;,.;]/i', ' ', $tweet);
$tweet = preg_replace('/rt | â€|/i', ' ', $tweet);
}

```

Gambar 4.6 Fungsi Proses Cleansing

c) *Convert Emoticon*

Proses *convert emoticon* ini akan ditangani oleh fungsi *convert_emoticon* kegunaannya untuk mengubah symbol *emoticon* pada *tweet* kedalam teks bahasa Indonesia. *Emoticon* ini dibagi menjadi dua, yaitu *emoticon* senang yang akan diubah menjadi kata ‘emotsenang’ dan *emoticon* sedih yang akan diubah menjadi kata ‘emotsedih’. *Emoticon* ini sangat berpengaruh terhadap hasil analisis sentimen, sehingga proses ini dapat dilihat pada gambar 4.7.

```

function convert_emoticon($tweet){
    $sesenang = array(">:",":-)",":o)","::]",":3)",":c)",":>","="),
    "8)",":)",":>)",":D)",":-D");
    $sesedih = array(">:[",":-((",":'(",":-c)",":c)",":-<",":-
    [",":{",":>.",":<.",":>.<",":/");
}

```

```

foreach ($sesenang as $item){
    $quotedSenang[] = preg_quote($item,'#');
}
$regexSenang = implode('|', $quotedSenang);
$fullSenang = '#(^|\W)('.$regexSenang.')($|\W)#';
foreach ($sesedih as $item){
    $quotedSedih[] = preg_quote($item,'#');
}
$regexSedih = implode('|', $quotedSedih);
$fullSedih = '#(^|\W)('.$regexSedih.')($|\W)#';
$tweet = preg_replace($fullSenang, ' emotsenang ', $tweet);
$tweet = preg_replace($fullSedih, ' emotsedih ', $tweet);
return $tweet;
}

```

Gambar 4.7 Fungsi Convert Emoticon

d) *Convert Negation*

Proses *convert negation* ini akan ditangani oleh fungsi *convert_negation* kegunaannya untuk mengubah artikata yang bersifat negasi dari kata – kata positif ke kedalam kata negatif dan juga sebaliknya. Jadi dalam fungsi *convert_negation* terdapat array list kata negasi, pada *tweet* yang mengandung kata negasi tersebut akan disambung dengan kata setelah kata negasi tersebut. Kode sumber *convert negation* dapat dilihat pada gambar 4.8.

```

function convert_negation($tweet){
    $list = array( 'gak ' => 'gak', 'ga ' => 'ga', 'ngga ' => 'ngga',
        'tidak ' => 'tidak', 'bkn ' => 'bkn', 'tida ' => 'tida', 'tak ' => 'tak',
        'jangan ' => 'jangan', 'enggak ' => 'enggak', 'gak ' => 'gak',
        'ga ' => 'ga', 'ngga ' => 'ngga', 'tidak ' => 'tidak',
        'bkn ' => 'bkn', 'tida ' => 'tida', 'tak ' => 'tak',
        'jangan ' => 'jangan', 'enggak ' => 'enggak'
    );
    $patterns = array();
    $replacement = array();
    foreach ($list as $from => $to) {

```

```

    $from = '/\b' . $from . '\b/';
    $patterns[] = $from;
    $replacement[] = $to;
}
return $tweet = preg_replace($patterns, $replacement, $tweet);
$tweet;
}

```

Gambar 4.8 Fungsi Convert Negation

e) Tokenizing

Proses *tokenizing* ini ditangani oleh fungsi *tokenizer* yang berguna untuk memisahkan setiap kata yang dihubungkan dengan karakter spasi menjadi setiap kata yang dihimpun pada array. Hasil dari proses tokenizer berupa data array yang dapat digunakan oleh proses selanjutnya. Fungsi *tokenizing* dapat dilihat pada gambar 4.9.

```

public function tokenizer($tweet){
    $tweet = stripslashes($tweet);
    $tweet = trim($tweet);
    $tweet = explode(" ", $tweet);
    return $tweet;
}

```

Gambar 4.9 Fungsi Tokenizing

f) Normalization

Proses *normalization* merupakan perubahan kata singkatan, kata inggris, dan kata yang tidak baku menjadi kata baku sesuai kamus bahasa Indonesia. Untuk menentukan suatu kata termasuk kata asing atau bukan didasarkan pada sebuah kumpulan kata pada *database*. Sebagai gambaran proses *normalization* dapat dilihat pada gambar 4.10.

```

public function normalization($tweet){
    $kata_tweet = $tweet;
    $i = 0;
    foreach ($kata_tweet as $kata_hasil) {
        $query_singkat = $this->con->query("SELECT * FROM tb_kata_singkatan
WHERE singkatan_kata = '". $kata_hasil. "'");
        if ($row = $query_singkat->fetch_array()) {
            $kata_tweet[$i] = $row[2];
        }
        $query_baku = $this->con->query("SELECT * FROM tb_kata_baku WHERE
baku_kata = '". $kata_hasil. "'");
        if ($row = $query_baku->fetch_array()) {
            $kata_tweet[$i] = $row[2];
        }
        $query_inggris = $this->con->query("SELECT * FROM tb_kata_inggris
WHERE inggris_kata = '". $kata_hasil. "'");
        if ($row = $query_inggris->fetch_array()) {
            $kata_tweet[$i] = $row[2];
        }
        $i++;
    }
    $kata = implode(' ', $kata_tweet);
    return $kata;
}

```

Gambar 4.10 Fungsi Normalization

g) *Stopword Removal*

Proses *Stopword Removal* ini digunakan untuk menghilangkan kata yang tidak berpengaruh dalam proses sentimen. Kata dari data *tweet* dibandingkan dengan kata yang terdapat dalam *database stopwords*, hasil dari proses ini adalah menghilangkan kata yang terdeteksi sama pada kata *stopword* yang ada pada *database*. Proses ini dapat digambarkan pada gambar 4.11.

```

function stopwords_removal($tweet){
    $stoplist = array();
    $query = $this->con->query("SELECT * FROM tb_kata_stopword");
    while ($key = $query->fetch_array()) {

```

```

    $stoplist[]= $key['stopword'];
  }
  $tweet = preg_replace('/\b(' . implode('|', $stoplist) . ')\b/', '', $tweet);
return $tweet;
}

```

Gambar 4.11 Fungsi Stopword Removal

h) Stemming

Proses terakhir dalam *preprocessing* adalah *stemming*, fungsi ini berguna untuk mengubah kata yang berimbuhan menjadi kata dasar agar memudahkan dalam proses pembobotan nantinya. Fungsi ini menggunakan algoritma Nazief Adriani dan penulis menggunakan library *stemming* yang dibuat oleh peneliti sebelumnya yaitu Sastrawi. Proses *stemming* dapat dilihat pada gambar 4.12.

```

function stemming($teks){
  $stemmerFactory = new \Sastrawi\Stemmer\StemmerFactory();
  $stemmer = $stemmerFactory->createStemmer();
  $sentence = $teks;
  $output = $stemmer->stem($sentence);
return $output;
}

```

Gambar 4.12 Fungsi Stemming

4.1.4 Pembobotan *Information Gain*

Proses pembobotan kata menggunakan algoritma *information gain*, bobot setiap kata akan digunakan untuk seleksi *fitur* kata yang memiliki bobot rendah. Proses ini dilakukan dengan cara menghitung bobot setiap kata yang ada dalam data *training*, kemudian perhitungan nilai *information gain* setiap kata akan dirangking untuk mendapatkan *fitur* yang terbaik menggunakan *threshold*. Pada penelitian kali ini penulis menggunakan 20 threshold yang berbeda untuk mengetahui tingkat akurasi yang dihasilkan dari perubahan threshold tersebut.

Akurasi terbaik yang didapatkan akan digunakan untuk menyeleksi *fitur* pada data *training*. Fungsi *information gain* dapat dilihat pada gambar 4.13.

```

public function entropy_set() {
    $jml_p = $this->get_jumlah_tweet_positif();
    $jml_n = $this->get_jumlah_tweet_negatif();
    $jml_t = $this->get_jumlah_tweet_total();
    $entropy_set = round(-(((($jml_p/$jml_t)*log($jml_p/$jml_t,2))+
($jml_n/$jml_t)*log($jml_n/$jml_t,2))),4);
    return $entropy_set;
}

public function set_info_gain() {
    $entropy_set = $this->entropy_set();
    $query = $this->con->query("SELECT `id_kata`, `entropy_kata` FROM
`data_training_kata` ORDER BY `id_kata` ASC");
    while ($row = $query->fetch_array()) {
        $sig = round($entropy_set - $row['entropy_kata'],9);
        $query_simpan = $this->con->query("UPDATE `data_training_kata` SET
`bobot_ig` = ".$sig." WHERE `id_kata` = ".$row['id_kata']."");
    }
}

```

Gambar 4.13 Fungsi Information Gain

4.1.5 Pembobotan *Naïve Bayes*

Dalam proses klasifikasi *Naïve Bayes* dibutuhkan bobot setiap kata yang ada dalam data *training*. Pembobotan ini dilakukan dengan cara menghitung probabilitas positif dan probabilitas negatif setiap kata yang ada dalam dokumen *training*. Implementasi dari perhitungan probabilitas setiap kata dilakukan dengan fungsi yang dibuat seperti gambar 4.14 untuk probabilitas positif dan gambar 4.15 untuk probabilitas negatif. Hasil dari pembobotan ini disimpan dalam *database* kata *training* yang selanjutnya akan digunakan bahan acuan untuk penentuan bobot data *testing*. Berikut potongan sumber kode yang digunakan untuk proses bobot kata data *training*.


```

function set_probabilitas_kata_positif() {
    $query = $this->con->query("SELECT `id_kata`, `kata` FROM
    `data_training_kata` ORDER BY `id_kata` ASC");
    while ($row_kata = $query->fetch_array()) {
        $ni = 0;
        $n = $this->get_jml_kata_positif();
        $kosakata = $this->get_jml_semua_kata_unik();
        $query_dokumen = $this->con->query("SELECT `id_training`,
        `tweet_preprocessing` FROM `data_training` WHERE `sentimen` =
        'P' ORDER BY `id_training` ASC");
        while ($row_dok = $query_dokumen->fetch_array()) {
            $kata_dok = explode(' ', $row_dok['tweet_preprocessing']);
            foreach ($kata_dok as $key) {
                if ($row_kata['kata'] == $key) {
                    $ni += 1;
                }
            }
        }
        $probabilitas_p = round(($ni+1)/($n+$kosakata),17);
        $query_simpan = $this->con->query("UPDATE `data_training_kata`
        SET `bobot_bayes_positif` = ".$probabilitas_p." WHERE
        `id_kata` = ".$row_kata['id_kata'].");
    }
}

```

Gambar 4.14 Fungsi Probabilitas Kata Positif

Fungsi *set_probabilitas_kata_positif* mengacu pada persamaan 2.11 pada bab II, oleh karena itu fungsi *set_probabilitas_kata_positif()* berfungsi untuk menghitung bobot setiap kata pada dokumen *tweet* data *testing*. Masukan yang dibutuhkan fungsi ini adalah jumlah kemunculan setiap kata pada kategori sentimen positif dan jumlah semua kata yang muncul pada kategori positif. Selain itu juga dibutuhkan masukan berupa jumlah dokumen pada kategori sentimen positif dan jumlah semua dokumen pada data training. Hasil dari fungsi ini dibuat berupa bobot kata positif yang digunakan untuk keperluan klasifikasi sentimen otomatis dengan metode *naïve bayes*.

```

set_probabilitas_kata_negatif() {
    $query = $this->con->query("SELECT `id_kata`, `kata` FROM
    `data_training_kata` ORDER BY `id_kata` ASC");
while ($row_kata = $query->fetch_array()) {
    $ni = 0;
    $n = $this->get_jml_kata_negatif();
    $kosakata = $this->get_jml_semua_kata_unik();
    $query_dokumen = $this->con->query("SELECT `id_training`,
    `tweet_preprocessing` FROM `data_training` WHERE `sentimen` =
    'N' ORDER BY `id_training` ASC");
while ($row_dok = $query_dokumen->fetch_array()) {
    $kata_dok = explode(' ', $row_dok['tweet_preprocessing']);
foreach ($kata_dok as $key) {
if ($row_kata['kata'] == $key) {
    $ni += 1;
    }
    }
}
    $probabilitas_p = round(($ni+1)/($n+$kosakata),17);
    $query_simpan = $this->con->query("UPDATE `data_training_kata`
    SET `bobot_bayes_negatif` = ".$probabilitas_p." WHERE
    `id_kata` = ".$row_kata['id_kata'].");
}
}

```

Gambar 4.15 Fungsi Probabilitas Kata Negatif

Fungsi `set_probabilitas_kata_negatif()` mengacu pada persamaan 2.11 pada bab II, oleh karena itu fungsi `set_probabilitas_kata_negatif()` berfungsi untuk menghitung bobot setiap kata pada dokumen *tweet* data *testing*. Masukan yang dibutuhkan fungsi ini adalah jumlah kemunculan setiap kata pada kategori sentimen negatif dan jumlah semua kata yang muncul pada kategori negatif. Selain itu juga dibutuhkan masukan berupa jumlah dokumen pada kategori sentimen negatif dan jumlah semua dokumen pada data training. Hasil dari fungsi ini dibuat berupa bobot kata negatif yang digunakan untuk keperluan klasifikasi sentimen otomatis dengan metode *naïve bayes*.

4.1.6 Klasifikasi Metode *Naïve Bayes*

Penentuan sentimen dilakukan dengan cara menghitung probabilitas dokumen data testing dengan merujuk pada probabilitas kata data training. Proses ini menggunakan algoritma *naïve bayes classifier*. Untuk implementasinya dapat dilihat pada gambar 4.16.

```
function klasifikasi_sentimen() {
    $query_dok = $this->con->query("SELECT `id_training`,
`tweet_preprocessing` FROM
data_training_tes ORDER BY `id_training`");
    while ($row_dok = $query_dok->fetch_array()) {
        $prob_kata_positif = []; $prob_kata_negatif = [];
        $kata_dok = $row_dok['tweet_preprocessing'];
        $kata_hasil = explode(' ', $kata_dok);
        foreach ($kata_hasil as $key) {
            $query_bobot_kata = $this->con->query("SELECT `id_kata`,
`kata`, `bobot_bayes_positif`, `bobot_bayes_negatif` FROM
`data_training_kata` WHERE `kata` = '". $key. "'");
            while ($row_kata = $query_bobot_kata->fetch_array()) {
                if ($key == $row_kata['kata']) {
                    $prob_kata_positif[$key] =
round($row_kata['bobot_bayes_positif'], 8);
                    $prob_kata_negatif[$key] =
round($row_kata['bobot_bayes_negatif'], 8);
                } else {
                    $prob_kata_positif[$key] = 1; $prob_kata_negatif[$key] = 1;
                }
            }
        }
        $prob_dokumen_positif =
$this->get_probabilitas_kategori_positif();
        foreach ($prob_kata_positif as $kata_prob => $value) {
            $prob_dokumen_positif *= $value;
        }
        $prob_dokumen_negatif =
$this->get_probabilitas_kategori_negatif();
        foreach ($prob_kata_negatif as $kata_prob => $value) {
            $prob_dokumen_negatif *= $value;
        }
    }
}
```

```
if ($prob_dokumen_positif > $prob_dokumen_negatif) {  
    $sentimen = "P";  
} else if ($prob_dokumen_positif < $prob_dokumen_negatif) {  
    $sentimen = "N";  
} else { $sentimen = "Tidak ada"; }  
}  
}
```

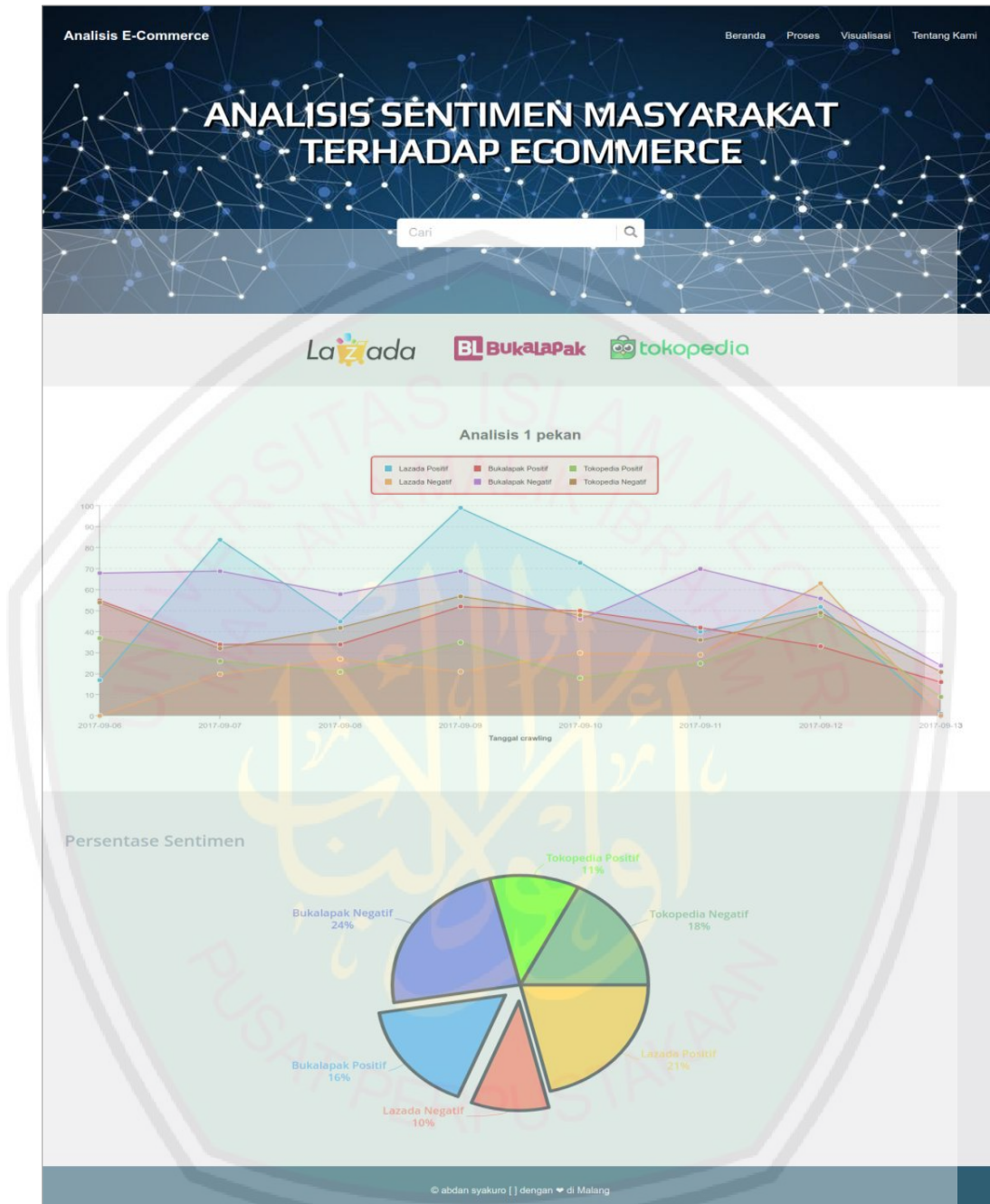
Gambar 4.16 Fungsi Naïve Bayes Classifier

Klasifikasi sentimen ini dilakukan secara otomatis dengan mengimplemmentasikan algoritma *Naïve Bayes Classification*. Proses ini diimplementasikan pada fungsi `klasifikasi_sentimen()` dengan membandingkan bobot setiap kata pada data *testing* dengan kata pada data *training*, jika kata tersebut tidak ditemukan dalam data *training* maka bobotnya dinilai 1. Hasilnya setiap dokumen training ini dijumlah bobot kata probabilitas positif dan probabilitas negatifnya. Selanjutnya bobot dokumen dibandingkan, jika bobot dokumen probabilitas positif lebih besar maka hasil sentimen adalah positif, dan jika bobot probabilitas negatif lebih besar maka hasil sentimen adalah negatif.

4.1.7 Desain dan Implementasi GUI

a) Halaman Utama

Pada halaman utama ini disajikan halaman menu header, form input pencarian, dan visualisasi. Menu header ini digunakan untuk mengakses halaman lainnya dalam aplikasi web. Kolom input pencarian digunakan untuk proses *crawling* suatu kata kunci yang langsung diproses samapai menampilkan sentimennya. Visualisasi pada halaman utama ini khusus untuk menampilkan data testing per satu pekan terakhir dihitung ketika terakhir proses *crawling*. Tampilan halaman utama bisa dilihat pada gambar 4.17.



Gambar 4.17 Tampilan Halaman Utama

b) Halaman *Crawling*

Pada halaman ini disajikan proses crawling dengan cara memasukkan kata kunci pada kolom pencarian yang telah disediakan. Berbeda dengan kolom pencarian halaman utama, pada halaman crawling ini gunanya hanya untuk proses crawling secara *realtime* dan hasilnya ditampilkan pada tabel

dibawahnya dan ditampung didalam database terlebih dahulu. Berikut tampilan proses crawling pada gambar 4.18.

Id Tweet	Tweet	Tanggal
907700997136441344	JAM TANGAN HUSH PUPPIES JAM TANGAN KU... https://t.co/ttHfJEi2i5	12-09-2017
907701131505057797	Serdadu Pelangi 2 https://t.co/54XidBn7i2	12-09-2017
907701338078715904	Sepatu MONNA VANIA 1688-21 https://t.co/FdYdZwVbWJ	12-09-2017
907702568595292160	Lembaran Cinta Prajurit https://t.co/mxVnvziAli	12-09-2017
907703684431151104	Belanja aman dan terpercaya check it now #diskon #ollshop #shopping	12-09-2017
907703952854024192	Wisata Maut https://t.co/f5ReGzuuxn	12-09-2017
907704294865911808	Belajar dan Pembelajaran - C. Asri Bu... https://t.co/F10Wzqm6WK	12-09-2017
907704331700363284	sepatu kerja untuk wanita https://t.co/kQ1PflQJc	12-09-2017

Gambar 4.18 Tampilan Proses Crawling

c) Halaman *Preprocessing*

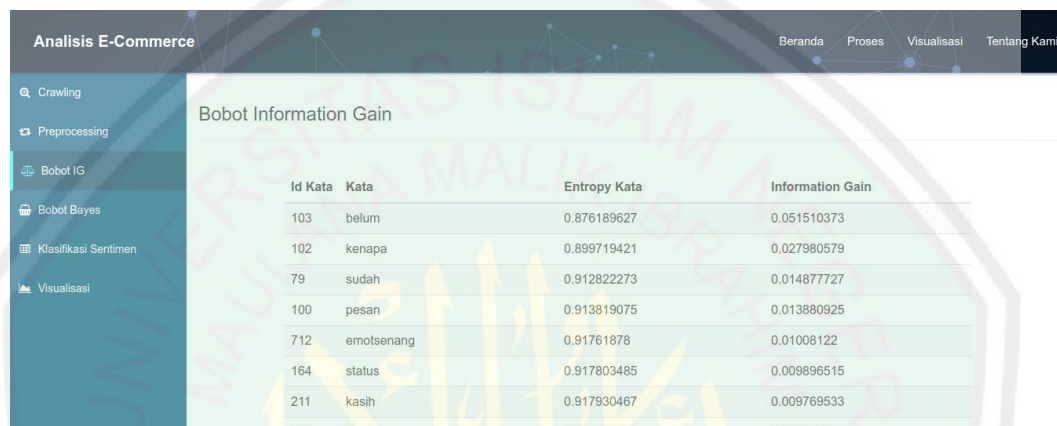
Pada halaman ini disajikan proses *preprocessing* dengan cara menekan tombol Proses. Maka ketika user menekan tombol proses, sistem akan mengolah data dari *database data crawling* yang hasilnya akan ditampung didalam *database data preprocessing* dan hasilnya bisa dilihat pada kolom dibawah tombol Proses. Tampilan halaman *preprocessing* dapat dilihat pada gambar 4.19.

Id Tweet	Tweet	Tweet Preprocessing	Tanggal
907700997136441344	JAM TANGAN HUSH PUPPIES JAM TANGAN KU... https://t.co/ttHfJEi2i5	hush puppies ku	12-09-2017
907701131505057797	Serdadu Pelangi 2 https://t.co/54XidBn7i2	serdadu pelangi	12-09-2017
907701338078715904	Sepatu MONNA VANIA 1688-21 https://t.co/FdYdZwVbWJ	sepatu monna vania cofdydzwbwj	12-09-2017
907702568595292160	Lembaran Cinta Prajurit https://t.co/mxVnvziAli	lembar cinta prajurit comxvnvziaili	12-09-2017
907703684431151104	Belanja aman dan terpercaya check it now #diskon #ollshop #shopping	belanja aman percaya periksa t cokmjewhglgi	12-09-2017
907703952854024192	Wisata Maut https://t.co/f5ReGzuuxn	wisata maut	12-09-2017
907704294865911808	Belajar dan Pembelajaran - C. Asri Bu... https://t.co/F10Wzqm6WK	ajar ajar asri bu	12-09-2017

Gambar 4.19 Tampilan Proses *Preprocessing*

d) Halaman Pembobotan Information Gain

Pada halaman ini disajikan tabel bobot *information gain* yang sudah diproses pada proses training. Halaman ini menampilkan kata dengan bobot *information gain* terbaik yang mengacu pada hasil akurasi. Tampilan bobot *information gain* dapat dilihat pada gambar 4.20 berikut.



Id Kata	Kata	Entropy Kata	Information Gain
103	belum	0.876189627	0.051510373
102	kenapa	0.899719421	0.027980579
79	sudah	0.912822273	0.014877727
100	pesan	0.913819075	0.013880925
712	emotsenang	0.91761878	0.01008122
164	status	0.917803485	0.009896515
211	kasih	0.917930467	0.009769533

Gambar 4.20 Tampilan Bobot *Information Gain*

e) Halaman Pembobotan Naïve Bayes

Halaman ini merupakan proses paling inti yaitu implementasi dari algoritma *Naïve Bayes*. Di halaman ini terdapat tombol Proses, yang ketika ditekan akan memproses perhitungan probabilitas positif dan probabilitas negatif dokumen *testing*. Hasil dari probabilitas ini dibandingkan mana yang lebih besar diantara kedua probabilitas tersebut, sehingga jika probabilitas positif yang lebih besar maka nilai sentimennya bisa diketahui yaitu Positif, dan sebaliknya jika probabilitas negatif lebih tinggi maka sentimen Negatif. Tampilan pembobotan *Naïve Bayes* dapat dilihat pada gambar 4.21.

Id	Tweet	Probabilitas Positif	Probabilitas Negatif
10099	JAM TANGAN HUSH PUPPIES JAM TANGAN KU... https://t.co/ttHfJE2i5	0.00003278051	0.00033911712
10098	Serdadu Pelangi 2 https://t.co/54XldBn7i2	0.343	0.657
10097	Sepatu MONNA VANIA 1688-21 https://t.co/FdYdZwVbWJ	0.00065557933	0.00033911712
10096	Lembaran Cinta Prajurit https://t.co/mxVnvzIAli	0.0000000783142775527	0.0000000070019546364
10095	Belanja aman dan terpercaya check it now https://t.co/kMjEWHgIGi ##diskon #ollshop #shopping	0.0000000000000061796270136348	0.0000000000000053722335006553

Gambar 4.21 Tampilan Proses Bobot Naïve Bayes

f) Halaman Klasifikasi Sentimen

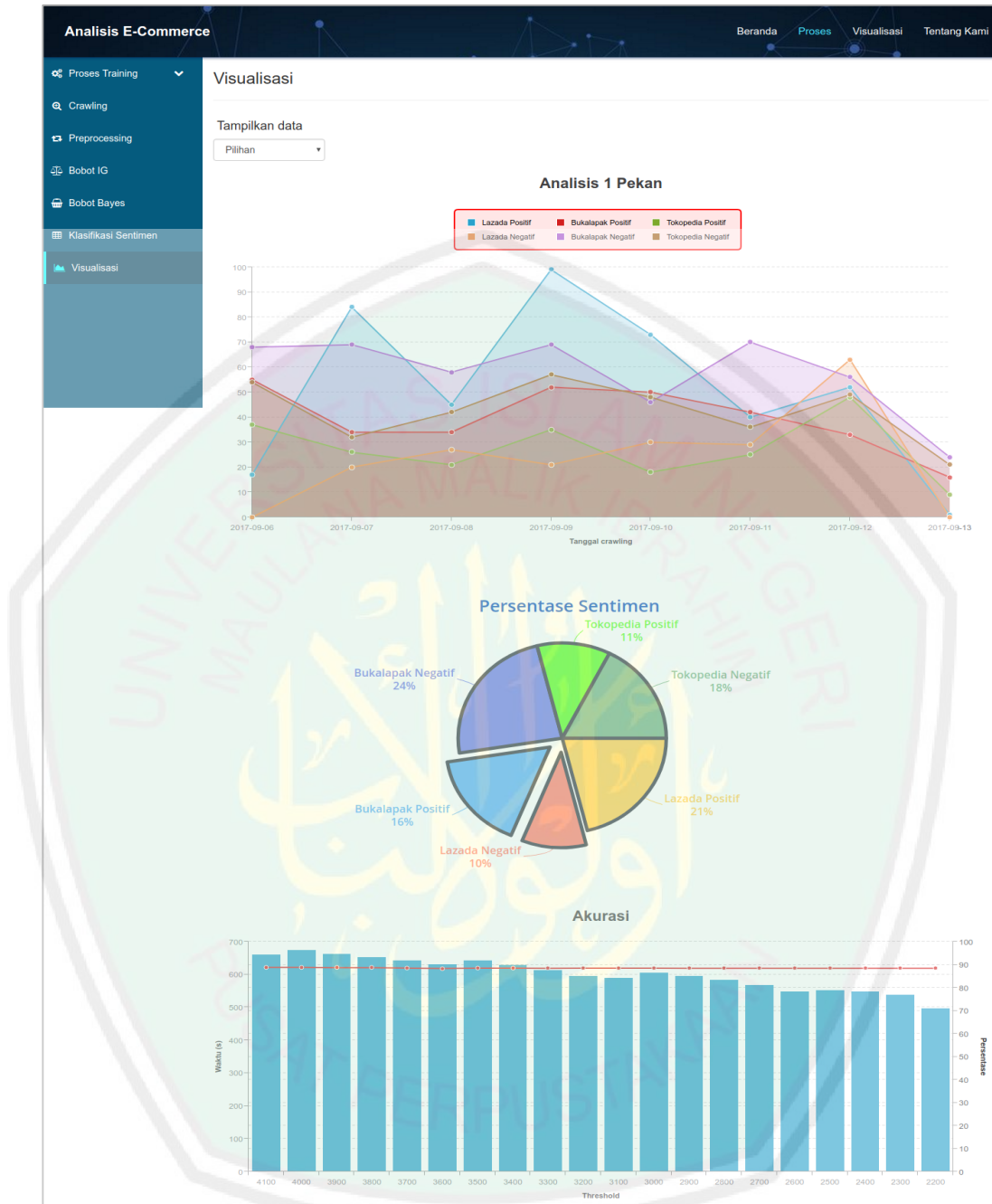
Halaman ini merupakan halaman hasil klasifikasi *naïve bayes*, dimana dihalaman ini ditampilkan dokumen testing serta sentimennya, yaitu positif atau negatif. Tampilan halaman klasifikasi sentimen dapat dilihat pada gambar 4.22 berikut.

Id	Tweet	Sentimen
10099	JAM TANGAN HUSH PUPPIES JAM TANGAN KU... https://t.co/ttHfJE2i5	Negatif
10098	Serdadu Pelangi 2 https://t.co/54XldBn7i2	Negatif
10097	Sepatu MONNA VANIA 1688-21 https://t.co/FdYdZwVbWJ	Positif
10096	Lembaran Cinta Prajurit https://t.co/mxVnvzIAli	Positif
10095	Belanja aman dan terpercaya check it now https://t.co/kMjEWHgIGi ##diskon #ollshop #shopping	Positif
10094	Wisata Maut https://t.co/f5ReGzuuxn	Negatif
10093	Belajar dan Pembelajaran - C. Asri Bu... https://t.co/F10Wzqm6WK	Positif
10092	sepatu kerja untuk wanita https://t.co/kQ1PfoJLc	Positif

Gambar 4.22 Tampilan Proses Klasifikasi Sentimen

g) Halaman Visualisasi

Halaman terakhir dari semua proses ini adalah visualisasi, pada penelitian kali ini penulis membuat 3 grafik visualisasi, yaitu grafik analisis sentimen perpekan, persentase perpekan, dan grafik akurasi metode *information gain*. Tampilan halaman visualisasi dapat dilihat pada gambar 4.23



Gambar 4.23 Tampilan Visualisasi

Pada gambar 4.23 terdapat tiga grafik, yang pertama menunjukkan jumlah sentimen setiap periode. Pada sistem visualisasi ini dibuat dengan 4 periode yaitu yang pertama tampilan data per 1 pekan, yang kedua tampilan data 1 bulan, yang ketiga tampilan data 6 bulan, dan yang terakhir tampilan data 1 tahun. Grafik yang

kedua menunjukkan jumlah sentimen dari semua periode yang dipilih. Pada grafik yang terakhir menunjukkan hasil akurasi sistem yang dibuat.

4.2 Pengujian dan Pembahasan Sistem

Pada sub bab ini akan dijelaskan mengenai analisis hasil pengujian sistem yang berfungsi untuk mengetahui kinerja program dalam proses klasifikasi. Pengujian ini dilakukan dengan cara menggunakan seluruh data training dijadikan sebagai data testing untuk menguji akurasi metode *naïve bayes* dengan seleksi *information gain*.

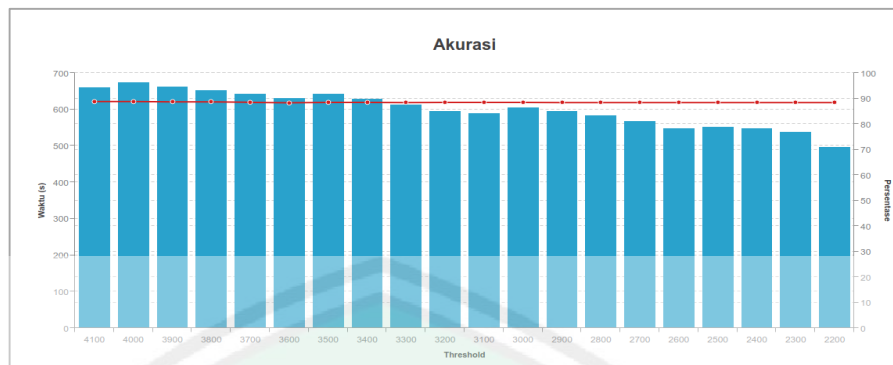
Evaluasi pengujian akurasi sistem dilakukan dengan memberi epoch = 20, sehingga hasil dari akurasi sistem ini berjumlah 20 kali percobaan. Pada penelitian ini setiap kali epoch bertambah diberikan threshold dengan cara mengurangi 100 kata pada setiap kata data training. Kata yang dikurangi diambil dari nilai bobot *information gain* yang paling rendah, sehingga menyisakan kata yang memiliki bobot terbaik. Hasil dari pengujian sistem ini dapat dilihat pada tabel 4.2 berikut.

Tabel 4.2 Hasil Akurasi

Epoch	Threshold	TP	TN	FP	FN	Akurasi	Waktu (s)
1	4100	733	1930	41	296	88.77 %	651.8943
2	4000	731	1933	38	298	88.80 %	672.8518
3	3900	728	1932	39	301	88.67 %	669.4441
4	3800	725	1935	36	304	88.67 %	646.3554
5	3700	720	1934	37	309	88.47 %	661.5116
6	3600	714	1936	35	315	88.33 %	629.6979
7	3500	716	1937	34	313	88.43 %	636.2398

8	3400	716	1937	34	313	88.43 %	644.808
9	3300	716	1936	35	313	88.40 %	629.2616
10	3200	716	1937	34	313	88.43 %	608.5906
11	3100	716	1937	34	313	88.43 %	599.5477
12	3000	716	1937	34	313	88.43 %	587.7442
13	2900	716	1936	35	313	88.40 %	591.497
14	2800	716	1936	35	313	88.40 %	564.8959
15	2700	716	1936	35	313	88.40 %	575.5903
16	2600	716	1936	35	313	88.40 %	568.4411
17	2500	716	1936	35	313	88.40 %	555.9261
18	2400	716	1936	35	313	88.40 %	536.6505
19	2300	716	1936	35	313	88.40 %	536.9626
20	2200	716	1936	35	313	88.40 %	514.0036

Hasil pengujian sistem pada tabel 4.2 bisa dilihat setiap pengurangan threshold membuat akurasi dan waktu proses berubah. Pada threshold 4000 didapatkan akurasi tertinggi yaitu 88.80 %, tetapi waktu prosesnya lebih lama dari sebelumnya. Pada pengurangan threshold berikutnya waktu proses relatif menurun, tetapi juga akurasi semakin menurun ketika threshold berkurang. Hal tersebut dikarenakan semakin threshold berkurang maka kata yang diolah semakin sedikit sehingga waktu bisa lebih cepat. Pada epoch 13 sampai 20 akurasi mengalami stagnan yaitu 88.40 %. Untuk lebih mudah memahami gambaran antara akurasi dengan kecepatan penulis membuat sebuah visualisasi grafik seperti pada gambar 4.24.



Gambar 4.24 Grafik Akurasi

Pada gambar 4.24 grafik akurasi digambarkan dengan *line* berwarna merah dan kecepatan proses digambarkan dengan diagram batang berwarna biru. Dapat diambil kesimpulan bahwa semakin berkurangnya *fiturinformation gain* kecepatan proses lebih cepat, dan akurasi hanya sedikit menurun, jadi jika sistem ini dibutuhkan untuk kecepatan maka bisa dipilih epoch ke-20. Pada penelitian ini penulis memilih akurasi tertinggi untuk bahan data *training* sehingga dipilih epoch ke 2.

4.3 Integrasi Penelitian dengan Islam

Penelitian yang dijalankan menghasilkan keluaran berupa sentimen masyarakat baik atau buruk terhadap suatu *e-commerce*, yaitu Lazada, Bukalapak, dan Tokopedia. Analisis sentimen masyarakat tersebut diharapkan akan menjadi tolak ukur suatu *e-commerce* dalam perbaikan pelayanannya, dan juga sebagai khazanah pengetahuan bagi masyarakat untuk menilai suatu produk *e-commerce*. Dengan demikian, keuntungan yang didapatkan adalah penyedia layanan *e-commerce* dapat melakukan perbaikan dari pendapat masyarakat terhadap produk yang dibuatnya. Untuk masyarakat bisa berhati – hati dalam memilih layanan *e-commerce* yang akan dia gunakan.

Allah ﷻ telah berfirman dalam Qur'an surah Hujurat ayat 6 bahwa sebagai manusia hendaklah memeriksa dahulu suatu berita yang datang kepadanya. Dengancara menanyakan kebenaran berita tersebut, karena jika berita tersebut merupakan berita tidak benar maka dapat mencelakakan sekumpulan manusia lainnya.

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنِّجَاءُ كُمْ فَاسِقِينَ بَاقْتَبَيْتُ أَنْ تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصِيبُوا
 حُوا عَلَّمَا فَعَلْتُمْ نَادِمِينَ (٦)

Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti, agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu (QS. Al-hujurat : 6)

Dalam tafsir fizhilalil qur'an disebutkan bahwa Allah memfokuskan orang fasik sebab dia dicurigai sebagai sumber kebohongan dan agar keraguan tidak menyebar di kalangan kaum muslimin karena berita yang disebarkan oleh setiap individunya, lalu ia menodai informasi. Pada prinsipnya, hendaklah setiap individu kaum muslimin menjadi sumber berita yang terpercaya dan hendaknya berita itu benar serta dapat dijadikan pegangan. Adapun orang fasik, maka dia menjadi sumber keraguan sehingga ini menjadi ketetapan. Dengan cara seperti itu, urusan umat menjadi lebih stabil dan moderat di antara mengambil dan menolak berita yang sampai kepadanya. Kaum muslimin jangan tergesa – gesa bertindak berdasarkan berita dari orang fasik. Pasalnya, ketergesa – gesaan itu membuatnya bertindak zalim kepada suatu kaum sehingga dia menyesal karena melakukan

perbuatan yang dimurkai Allah serta tidak mempertahankan kebenaran dan keadilan.

Ayat tersebut telah memotivasi penulis untuk mencari solusi terhadap masalah sentimen masyarakat terhadap *e-commerce*. Banyaknya pengguna *e-commerce* dan banyaknya pendapat masyarakat dalam menanggapi pelayanan dari pihak *e-commerce* membuat bingung jika harus menanggapi satu persatu pendapat masyarakat. Akhirnya solusi yang diperoleh untuk mempermudah analisis sentimen masyarakat terhadap *e-commerce* menggunakan metode *Naïve Bayes* dengan seleksi fitur *Information Gain*. Seseuai dengan firman Allah ﷻ, bahwa manusia harus melakukan *tabayyun* jika mendapatkan suatu berita dari orang lain.



BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilaksanakan, dapat disimpulkan bahwa:

1. Metode *Naïve Bayes* dengan seleksi fitur *Information Gain* terbukti dapat menganalisis sentimen secara otomatis. Uji coba dilakukan dengan menggunakan data *testing* secara *realtime*, setiap data diklasifikasikan dengan sentimen positif atau negatif.
2. Performa dari penggabungan metode *Naïve Bayes* dengan seleksi fitur *Information Gain* mengalami peningkatan yaitu pada batas *threshold* 2200 waktu proses menjadi 514 detik lebih cepat dari sebelum penambahan *Information Gain*. Sistem analisis sentimen jika ditambah dengan seleksi fitur *Information Gain* dapat menambah akurasi hingga mencapai 88.8%.

5.2 Saran

Beberapa saran untuk pengembangan penelitian dimasa akan datang adalah sebagai berikut :

1. Pada penelitian selanjutnya penulis menyarankan menambahkan metode *Adaboost* untuk mengurangi bias agar dapat meningkatkan akurasi algoritma *Naïve Bayes* secara signifikan.
2. Penambahan koleksi kamus pada database kata gaul, karena pada sosial media *twitter* terlalu banyak bahasa yang kurang baku.

DAFTAR PUSTAKA

- Aditya, B. R. (2015). Penggunaan Web Crawler untuk Menghimpun Tweet dengan Metode Pre-Processing Text Mining. *Jurnal Infotel Vol. 7 No. 2* .
- Aditya, C. S., Hani'ah, M., Fitrawan, A. A., Arifin, A. Z., & Purwitasari, D. (2016). Deteksi Bot Spammer pada Twitter Berbasis Sentiment Analysis. 7.
- Agusta, L. (2009). Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. *Konferensi Nasional Sistem dan Informatika (KNSN)* .
- Chandani, V., Wahono, R. S., & Purwanto. (2015). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature. *Journal of Intelligent Systems* .
- Faradhillah, N. Y., Kusumawardani, R. P., & Hafidz, I. (2016). Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin. *Seminar Nasional Sistem Informasi Indonesia* .
- Hadna, N. M., Santosa, P. I., & Winarno, W. W. (2016). Studi Literatur tentang Perbandingan Metode untuk Proses Analisis Sentimen di Twitter. *Seminar Teknologi Informasi dan komunikasi (SENTIKA)* .
- Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Penerbit Andi.
- Kurniawan, B., Effendi, S., & Sitompul, O. S. (2012). Klasifikasi Konten Berita Dengan Metode Text Mining. *JURNAL DUNIA TEKNOLOGI INFORMASI Vol. 1, No. 1* , 14 - 19.
- Maulida, I., Suyatno, A., & Hatta, H. R. (2016). Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain. *JSM STM IK* .
- Mujilawati, S. (2016). Pre-processing Text Mining pada Twitter. *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)* .
- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *Program Studi Teknik Informatika, Fakultas Informatika, Universitas Telkom, Bandung* .
- Nurhuda, F., Sihwi, S. W., & Doewes, A. (2013). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter menggunakan Metode Naive Bayes Classifier. *Jurnal ITSMART* .

- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* , 6.
- Putranti, D. N., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa. *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)* .
- Rasywir, E., & Purwarianti, A. (2015). Eksperimen pada Sistem Klasifikasi Berita Hoax Bebahasa Indonesia Berbasis Pembelajaran Mesin. *Jurnal Cybermatika* .
- Rodiyansyah, S. F., & Winarko, E. (2013). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)* .
- Rozi, I. F., Pranomo, S. H., & Dahlan, E. A. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Electrics, Electronics, Communications, Controls, Informatics, Systems (EECCIS)* .
- Saadah, M. N., Atmagi, R. W., Rahayu, D. S., & Arifin, A. Z. (2013). Sistem Temu Kembali Dokumen Teks dengan Pembobotan TF-IDF dan LCS. *Jurnal Ilmiah Teknologi Informasi (JUTI)* .
- Setyani, N. E. (2013). Penggunaan Media Sosial Sebagai Sarana Komunikasi Bagi Komunitas. *Jurnal Komunikasi Universitas Sebelas Maret* .
- Sutojo, T., Mulyanto, E., Suhartono, & Vincent. (2011). *Kecerdasan Buatan*. Yogyakarta: ANDI.
- Thumbs up? Sentiment Classification using Machine Learning 2002 *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* Philadelphia Association for Computational Linguistics
- Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan menggunakan Naive Bayes. *Jurnal Evolusi* .