

**ANALYSIS SENTIMEN PADA MEDIA SOSIAL  
MENGUNAKAN METODE NAIVE BAYES CLASSIFIER**

**TESIS**

**Oleh:**

**NUARI ANISA SIVI  
NPM. 1521210039**



**MAGISTER TEK INFORMATIKA  
PROGRAM PASCA SARJANA  
INFORMATICS & BUSINESS INSTITUTE DARMAJAYA  
BANDAR LAMPUNG  
2018**

## PERNYATAAN KEASLIAN LAPORAN TESIS

Saya yang bertandatangan dibawah ini adalah :

**Nama** : Nuari Anisa Sivi  
**NPM** : 1521210039  
**Program Studi** : Magister Teknik Informatika  
**Judul Tesis** : ANALYSIS SENTIMEN PADA MEDIA SOSIAL  
MENGUNAKAN METODE NAIVE BAYES  
CLASSIFIER

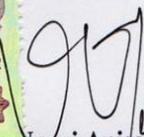
Dengan ini menyatakan bahwa :

1. Tesis ini adalah asli dan bukan plagiat sebagai atau keseluruhan, serta belum pernah diajukan untuk mendapatkan gelar akademik dimanapun.
2. Karya tulis ini adalah murni gagasan, rumusan, dan penelitian saya sendiri melalui proses pembimbingan dan pengujian.
3. Semua kutipan, petikan dari karya orang lain saya tuliskan namanya dengan jelas serta saya cantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sebenar-benarnya, jika terdapat penyimpangan di kemudian hari, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma dan aturan yang berlaku.

Bandar Lampung, 1 Februari 2018

Yang menyatakan,



  
**Nuari Anisa Sivi**  
**NPM.1521210039**

**LEMBAR PERSETUJUAN TESIS**

Judul Tesis : **ANALYSIS SENTIMEN PADA MEDIA SOSIAL  
MENGUNAKAN METODE NAIVE BAYES  
CLASSIFIER**

Nama Mahasiswa : **NUARI ANISA SIVI**

NPM : 1521211039

Program Studi : Magister Teknik Informatika

Telah diperiksa dan disetujui untuk diajukan dalam ujian sidang tesis guna memperoleh gelar Magister Teknik Informatika pada program pascasarjana IIB Darmajaya

Bandar Lampung, 1 Februari 2018

**Menyetujui,**

Dosen Pembimbing 1

Dosen Pembimbing 2

**R.Z. Abdul Aziz, Ph.D**

NIK 01050904

**Fitria, S.T., M.Kom**

NIK 00490802

**Mengetahui,**

Ka. Prodi Magister Teknik Informatika

**Dr. Sutedi, S.Kom., M.T.I**

NIK 00600303

## LEMBAR PENGESAHAN TESIS

Pada hari Kamis, tanggal 1 Februari 2018, telah diselenggarakan ujian sidang Tesis dengan judul **ANALYSIS SENTIMEN PADA MEDIA SOSIAL MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER**. Untuk memenuhi persyaratan akademik guna memperoleh gelar Magister Teknik Informatika pada Program Pascasarjana Informatics and Business Institute Darmajaya, bagi mahasiswa:

Nama Mahasiswa : **NUARI ANISA SIVI**

NPM : 1521211039

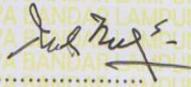
Program Studi : Magister Teknik Informatika

Telah dinyatakan Lulus oleh Dewan Penguji yang terdiri dari :

Nama Tim Penguji Tanda Tangan

1. **Dr. Kurnia Muludi S.**

Ketua Sidang



2. **Sriyanto, S.Kom., M.M., Ph.D**

Anggota



Dekan Fakultas Ilmu Komputer  
Program Studi Magister Teknik Informatika



**Zaidir Jamal, S.T., M.E.Eng**

NIK 00590203

## **MOTTO**

*”Ilmu tanpa Agama adalah BUTA,  
Agama tanpa ilmu adalah SESAT”*

## PERSEMBAHAN

Dengan mengucap rasa syukur dan bahagia kepada Allah SWT atas seluruh nikmat yang di berikan. Shalawat serta salam yang selalu tercurah kepada Nabi besar Muhammad SAW yang telah mensyiarkan kemuliaan Islam ke seluruh penjuru dunia hingga penulis rasakan sekarang. Penulis persembahkan karya ini kepada orang-orang yang telah banyak membantu dan memotivasi dalam pembuatannya :

- Terima kasihku kepada orang tua ku tercinta, *Ayahanda Sujarni* dan *Ibunda Sariyem (Nung)* atas didikan, pengorbanan dan darah yang mengalir dalam tubuhku. Serta saudaraku yang senantiasa memberi dukungan dan mensupport baik dari segi moril ataupun materil.
- Anakku *Alwi Assegaf* Dan *Abidzar Ar- Rasyid*, yang selalu memberikan motivasi agar tetap semangat dalam menjalani hidup ini.
- Ibu *Fitria*. Selaku Pembimbing penulisan Tesis dan Dosen - dosen lainnya yang selalu memberikan motivasi dan memberikan semangat dalam penyelesaian Tesis ini.
- Almamaterku yang menjadikan pola pikirku maju.
- serta teman - temanku yang tidak dapat kusebutkan satu persatu yang selalu memberikan motivasi agar tetap semangat dalam menjalani pendidikan ini.

## KATA PENGANTAR

Dengan memanjatkan puji dan syukur ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya pada penulis, akhirnya penulis dapat menyelesaikan penyusunan tesis yang berjudul **Analysis Sentimen Pada Media Sosial Menggunakan Metode Naive Bayes Classifier** Tesis ini ditulis dalam rangka memenuhi salah satu persyaratan memperoleh gelar Magister Teknik Informatika di Program Pascasarjana IBI Darmajaya Lampung. Penulis menyadari bahwa tesis ini dapat diselesaikan berkat dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis berterimakasih kepada semua pihak yang secara langsung dan tidak langsung memberikan kontribusi dalam penyelesaian tesis ini. Secara khusus pada kesempatan ini penulis menyampaikan terimakasih kepada:

1. Bapak Ir. Firmansyah YA, M.B.A., M.Sc. selaku Rektor IBI Darmajaya.
2. Bapak Zaidir Jamal, ST., M.E.Eng, selaku Dekan Fakultas Ilmu Komputer Program Studi Magister Teknik Informatika IBI Darmajaya;
3. Bapak Dr. Sutedi, S.Kom.,M.T.I, selaku ketua program studi Magister Teknik Informatika IBI Darmajaya
4. Bapak R.Z. Abdul Aziz, Ph.D (pembimbing 1) dan ibu Fitria, ST., M.Kom (pembimbing II) selaku pembimbing penulisan Tesis.
5. Kedua orang tua tercinta. Bapak *Sujarni* dan Ibu *Sariyem (Nung)* yang telah mendidik dan membesarkan penulis, mendoakan dan memberikan cinta serta kasih sayang yang tak terhingga.
6. Para dosen Pascasarjana IBI Darmajaya Lampung yang telah membimbing dengan ilmu-ilmu yang bermanfaat dan memotivasi penulis dalam penyelesaian tesis ini.

Penulis berharap tesis ini mudah-mudahan dapat memberikan sumbangsih bagi pendidikan yang selalu menghadapi tantangan seiring dengan tuntutan zaman.

Bandar Lampung, 1 Februari 2018

**Nuari Anisa Sivi**  
**NPM. 1521210039**

# ANALYSIS SENTIMEN PADA MEDIA SOSIAL MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

NUARI ANISA SIVI

## ABSTRAK

Media sosial *Facebook* dan *Twitter* merupakan salah satu contoh media sosial yang digunakan masyarakat untuk saling berinteraksi satu sama lain. *Facebook* dan *Twitter* memberikan layanan kepada penggunanya untuk mengirim dan membaca *Facebook* dan *Tweets* yang telah dibagikan, sehingga masyarakat lebih memilih menuangkan opininya melalui media sosial daripada meenyampaikan secara langsung. Opini masyarakat yang tertuang dalam media sosial berupa sebuah persepsi, baik itu positif maupun negatif. Melimpahnya opini masyarakat dapat dimanfaatkan sebagai bahan penelitian untuk mencari sebuah informasi. Pemanfaatan informasi tersebut membutuhkan teknik analisis yang tepat sehingga informasi yang dihasilkan mampu membantu banyak pihak dalam mengambil sebuah keputusan. Penggunaan teknik dalam pengolahan data dapat diselesaikan dengan teknik analisa sentimen atau opinion mining.

Oleh sebab itu, pada penelitian ini mencoba melakukan analisa sentimen untuk melihat persepsi masyarakat terhadap para pejabat Negara ataupun Gubernur yang sedang menjalani masa jabatannya menggunakan *Metode Naïve Bayes Classifier* dengan mengkalsifikasikan sentimen menjadi positif, negatif dan netral.

Penulisan ini menggunakan *Tools weka 3.9.2* untuk melihat seberapa akuratnya *Metode Naïve Bayes Classifier* dengan menggunakan Data Positif 100, Negative 100 dan Netral 100 dengan jumlah kata yang sering keluar pada rentang waktu 6 bulan.

Hasil penelitian yang telah dilakukan dapat diketahui bahwa *Metode Naïve Bayes Classifier* menunjukkan hasil akurasi data yang besar yaitu bisa mencapai 99% dalam menentukan kata sentimen.

Selain itu, hasil dari pengujian kinerja sistem menggunakan data latihan 10 positif, 10 negatif dan 10 netral dengan *Metode Naïve Bayes Classifier* menghasilkan nilai akurasi klasifikasi yang lebih baik.

**Kata kunci :** *Analisis sentymen, Naive bayes classifier, Media Social, Pemimpin*

## SENTIMENT ANALYSIS OF SOCIAL MEDIA USING THE *NAIVE BAYES* CLASSIFIER METHOD

By:  
NUARI ANISA SIVI

### ABSTRACT

Facebook and Twitter social media are examples of social media that people use to interact with one another. Facebook and Twitter provide services to users to post and read Facebook and Tweets that have been shared, so people prefer to express their opinions through social media rather than conveying it directly. Public opinion expressed in social media in the form of a perception can be both positive and negative.

The abundance of public opinion can be used as research materials to find information. The use of this information requires the right analytical techniques so that the information generated is able to help many parties in making a decision. The use of techniques in data processing can be solved by sentiment analysis or opinion mining techniques. Therefore, this research tried to do the sentiment analysis to see public perceptions of State officials or governors who are undergoing their tenure using the *Naïve Bayes Classifier* method by calcifying sentiments to be positive, negative and neutral.

This study used Weka 3.9.2 tools to see how accurate the *Naïve Bayes Classifier* method was by using 100 positive, 100 negative and 100 neutral data with the number of words that often come out in a span of 6 months.

The results of the research that has been done can be seen that the *Naïve Bayes Classifier* method showed the results of large data accuracy that reached 99% in determining the sentiment words.

In addition, the results of system performance testing using 10 positive, 10 negative and 10 neutral training data with the *Naïve Bayes Classifier* method produced better classification accuracy values.

**Keywords:** sentiment analysis, *naive bayes classifier*, social media, leaders



## DAFTAR RIWAYAT HIDUP

### Data Pribadi

Nama : NUARI ANISA SIVI  
Tempat, Tanggal Lahir : Balerejo, 28 Januari 1991  
Jenis Kelamin : Perempuan  
Agama : Islam  
Kewarganegaraan : Indonesia  
Alamat : Dusun gunantoro, Rt/Rw.010/003, Desa Balerejo,  
Kec Btang Hari  
E-mail : [nuarianisasivi@gmail.com](mailto:nuarianisasivi@gmail.com)

### Riwayat Pendidikan

- ✚ Pendidikan Sekolah Dasar (SD) Negeri 01 Balerejo Lampung Timur, lulus tahun 2002
- ✚ Madrasah Tsanawiyah (Mts) Negeri 01 Lampung timur, lulus tahun 2005.
- ✚ Sekolah Menengah Kejuruan (SMK) Negeri 3 Metro, lulus tahun 2008
- ✚ S1 Teknik Informatika di Informatics and Busines Institute (IBI) Darmajaya Bandar Lampung, lulus Tahun 2013
- ✚ Pada tahun 2015 Penulis diterima di IBI Darmajaya Jurusan S2 Magister Teknik Informatika.

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>LEMBAR PERNYATAAN KEASLIAN LAPORAN TESIS .....</b>	<b>ii</b>
<b>LEMBAR PERSETUJUAN TESIS.....</b>	<b>iii</b>
<b>LEMBAR PENGESAHAN SIDANG TESIS .....</b>	<b>iv</b>
<b>MOTTO .....</b>	<b>v</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>ABSTRAK .....</b>	<b>ix</b>
<b>ABSTRACT .....</b>	<b>x</b>
<b>DAFTAR RIWAYAT HIDUP .....</b>	<b>xi</b>
<b>DAFTAR ISI.....</b>	<b>xii</b>
<b>DAFTAR TABEL .....</b>	<b>xiii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	2
1.3 Ruang Lingkup.....	2
1.4 Tujuan Penelitian .....	2
1.5 Manfaat Penelitian .....	3
1.6 Sistematika Penulisan .....	3
<b>BAB II LANDASAN TEORI .....</b>	<b>5</b>
2.1 <i>Data Mining</i> .....	5
2.1.1 Pengertian Data Mining.....	5
2.1.2 Fungsi Data Mining .....	5
2.1.3 Proses Data Mining.....	5
2.1.4 Klasifikasi .....	7
2.2 <i>Text Mining</i> .....	8

2.3 Teknik Evaluasi Dalam Data Mining .....	8
2.4 Sentiment Analysis .....	10
2.5 Preprocessing .....	10
2.6 Pengertian Social Media .....	14
2.6.1 Karakteristik Social Media .....	14
2.6.2 Jenis Social Media .....	15
2.6.3 Pengertian <i>Twitter</i> .....	16
2.6.4 Facebook.....	16
2.7 Algoritma Naive Bayes Classifier .....	16
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>20</b>
3.1 Metode Penelitian .....	20
3.1.1 Metode Pengumpulan Data.....	20
3.2 Implementasi .....	20
3.1.3 Metode Pengembangan Perangkat Lunak.....	22
3.1.3.1 Perencanaan .....	22
3.1.3.2 Analisis .....	22
3.1.3.3 Pembuatan.....	23
1. Mengubah File CSV menjadi ARF.....	23
2. <i>Pre-processing</i> (Persiapan data) .....	25
3.3 <i>Evaluasi Hasil</i> .....	30
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>31</b>
4.1 Implementasi.....	31
4.1.1 Skenario Pengujian .....	31
4.2 Analisis Sistem.....	36
<b>BAB V SIMPULAN DAN SARAN .....</b>	<b>38</b>
5.1 Simpulan .....	38
5.2 Saran .....	38

## DAFTAR PUSTAKA

## LAMPIRAN

## DAFTAR TABEL

Tabel 2.3 Model <i>Confusion Matrix</i> .....	9
Tabel 2.5 Daftar <i>Convert Emoticon</i> .....	11
Tabel 3.1 Tabel Data Shet Sentimen .....	21
Tabel 3.2 Contoh Daftar Kata Sentimen .....	35
Tabel <b>3.3</b> Confusion Matrik .....	30
Tabel <b>4.1 Jumlah Data</b> .....	32
Tabel <b>4.2 Summary</b> .....	32
Tabel <b>4.3 Detailed Accuracy By Class</b> .....	32
Tabel <b>4.4 Confusion Matrix</b> .....	32
Tabel <b>4.5 Jumlah Data</b> .....	33
Tabel <b>4.6 Summary</b> .....	33
Tabel <b>4.7 Detailed Accuracy By Class</b> .....	33
Tabel <b>4.8 Confusion Matrix</b> .....	34
Tabel <b>4.9 Jumlah Data</b> .....	34
Tabel <b>4.10 Summary</b> .....	34
Tabel <b>4.11 Detailed Accuracy By Class</b> .....	35
Tabel <b>4.12 Confusion Matrix</b> .....	35

## DAFTAR GAMBAR

Gambar 2.1 Teknik Klasifikasi .....	8
Gambar <b>2.5.1 Contoh Case Folding</b> .....	11
Gambar 2.5.2 Contoh <i>Cleansing</i> .....	12
Gambar. 2.5.3 Contoh <i>Tokenizing</i> .....	12
Gambar 2.5.4 Contoh <i>Stopword Removal</i> .....	13
Gambar 2.5.5 Contoh <i>Stemming</i> .....	13
Gambar 3.2 Diagram Alir Kerja Penelitian.....	21
Gambar <b>3.2 Gambar Aplikasi WEKA</b> .....	23
Gambar <b>3.3 Gambar Open Exploler pada Aplikasi WEKA</b> .....	24
Gambar. <b>3.4 Gambar Open File pada Aplikasi WEKA</b> .....	24
Gambar. <b>3.5 Gambar Mengubah format datshet ke ARF</b> .....	25
Gambar <b>3.7. TF-IDF (<i>Term Frequency – Inverse Document Frequency</i>)</b> .....	29
Gambar. <b>3.8. Ilustrasi 10 fold cross validation</b> .....	30

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Media jejaring sosial seperti Twitter, Facebook dan Youtube merupakan beberapa media perangkat komunikasi terpopuler di masyarakat saat ini . Media sosial telah menjadi *trend* dari kehidupan masyarakat luas dan menjadi identitas mereka dalam bersosialisasi, tidak hanya bagi kalangan sendiri, tetapi juga bagi masyarakat luas dan dunia. Facebook merupakan situs media sosial yang memiliki pengguna aktif terbanyak di dunia. Sejak berdirinya Facebook pada tahun 2004 hingga Juni 2014, Facebook telah memiliki 1,32 Milyar pengguna aktif. Dimana Indonesia menduduki peringkat ke empat dengan jumlah 65 juta pengguna Facebook aktif.

Informasi yang terdapat didalam facebook mayoritas adalah berupa data teks. Opini serta pendapat tertulis dalam kalimat-kalimat yang berupa teks dan tidak terstruktur. Oleh karena itu diperlukan sebuah analisis yang tepat untuk mengolah data teks sehingga memberikan informasi yang akurat. Analisis sentimen akan memberikan suatu pengambilan informasi berkualitas tinggi dari teks untuk pengklasifikasian atau kategorisasi.

Analisis sentimen adalah studi komputasional dari opini-opini orang, penilaian dan emosi melalui entitas, *event* dan atribut yang dimiliki. Tugas dasar analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau fitur/tingkat aspek apakah pendapat yang dikemukakan dalam teks tersebut bersifat positif, negatif atau netral .

*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 yang menarik dari sekumpulan data teks yang berjumlah besar . Salah satu metode klasif

ikasi yang dapat digunakan dalam melakukan *text mining* adalah metode *Naive Bayes Classifier* (NBC). Kelebihan NBC yaitu merupakan metode yang sederhana tetapi memiliki tingkat akurasi tinggi. NBC terbukti dapat digunakan secara efektif untuk mengklasifikasikan teks secara otomatis. Algoritma NBC yang sederhana dan kecepatannya tinggi dalam proses pelatihan dan klasifikasi membuat algoritma ini menarik untuk digunakan sebagai salah satu metode klasifikasi .

Berdasarkan latar belakang yang telah dijabarkan, maka dalam penelitian ini mengangkat judul proposal yang berjudul “Penggunaan Metode Metode *Naive Bayes Classifier* (NBC) Untuk Menganalisis Sentimen Pada Sosial Media”.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang diatas, maka rumusan masalah untuk penelitian ini adalah Bagaimana mengimplementasikan metode *Naive Bayes Classifier* untuk analisis sentimen positif, negatif dan netral?

## **1.3 Batasan Masalah**

Agar pembahasan dalam penelitian ini dapat lebih terarah maka penulisan memberikan batasan masalah yaitu :

1. Kategori sentimen yang digunakan meliputi sentimen positif, negatif dan netral.
2. Data yang digunakan berupa teks, tidak menggunakan simbol, angka, gambar dan vidio.
3. Data opini yang digunakan merupakan opini bahasa Indonesia.
4. Sumber data didapat melalui grup Facebook Muhammad Ridho Ficardo Fans.

## **1.4 Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah :

1. Mengklasifikasikan sentimen pada opini Facebook dengan proses *Text Mining* menggunakan metode *Naive Bayes Classifier* (NBC).

2. Melakukan analisis terhadap hasil perhitungan metode *Naive Bayes Classifier* (NBC).
3. Mengetahui hasil akurasi yang optimal menggunakan metode *Naive Bayes Classifier* (NBC).

### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah :

1. Mempermudah proses pengklasifikasian opini Facebook berbahasa Indonesia.
2. Mengetahui tingkat akurasi dalam pengklasifikasian teks menggunakan *Naive Bayes Classifier* (NBC).
3. Memberikan informasi secara cepat tentang reaksi (sentimen) pembaca terhadap kebijakan Institusi.

### **1.6 Sistematika Penulisan**

#### **BAB I            PENDAHULUAN**

Bab ini menjelaskan Latar Belakang, Perumusan Masalah, Batasan Masalah, Tujuan dan Manfaat dan Sistematika Penulisan.

#### **BAB II           LANDASAN TEORI**

Bab ini membahas tentang tinjauan terhadap objek yang diteliti dan berisi tentang teori-teori tentang sistem pakar, serta metode-metode yang digunakan yang berkaitan dengan topik penelitian.

#### **BAB III          METODE PENELITIAN**

Menjelaskan diagram alir pemecahan masalah, metode pengumpulan data, metode pengembangan perangkat lunak, analisis sistem yang di usulkan, desain global sistem yang diusulkan, konteks diagram, rancangan *database*, dan rancangan *interface*.

#### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini menyajikan hasil tampilan program, penjelasan dan pembahasan.

#### **BAB V SIMPULAN DAN SARAN**

Bab ini berisi kesimpulan pembahasan tentang hasil yang telah diperoleh dan saran-saran yang memungkinkan untuk pengembangan skripsi ini.

#### **DAFTAR PUSTAKA**

#### **LAMPIRAN**

## **BAB II LANDASAN TEORI**

### **2.1 Data Mining**

#### **2.1.1 Pengertian Data Mining**

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. (Turban, dkk. 2005).

#### **2.1.2 Fungsi Data Mining**

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Pada dasarnya, data mining mempunyai empat fungsi dasar yaitu : (Turban, dkk. 2005).

1. Fungsi Klasifikasi (Classification)

Data mining dapat digunakan untuk mengelompokkan data-data yang jumlahnya besar menjadi data-data yang lebih kecil.

2. Fungsi Segmentasi (Segmentation)

Data mining dapat digunakan untuk melakukan segmentasi (pembagian) terhadap data berdasarkan karakteristik tertentu.

3. Fungsi Asosiasi (Association)

Data mining digunakan untuk mencari hubungan antara karakteristik tertentu.

4. Fungsi Pengurutan (Sequencing)

Data mining digunakan untuk mengidentifikasi perubahan pola yang telah terjadi dalam jangka waktu tertentu.

#### **2.1.3 Proses Data Mining**

Proses dalam data mining ada 6 yaitu : (Turban, dkk. 2005).

1. Pembersihan data (Data Cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh,

baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

## 2. Integrasi Data (Data Integration)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

## 3. Transformasi Data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

## 4. Proses Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

#### 5. Evaluasi Pola (Pattern Evaluation)

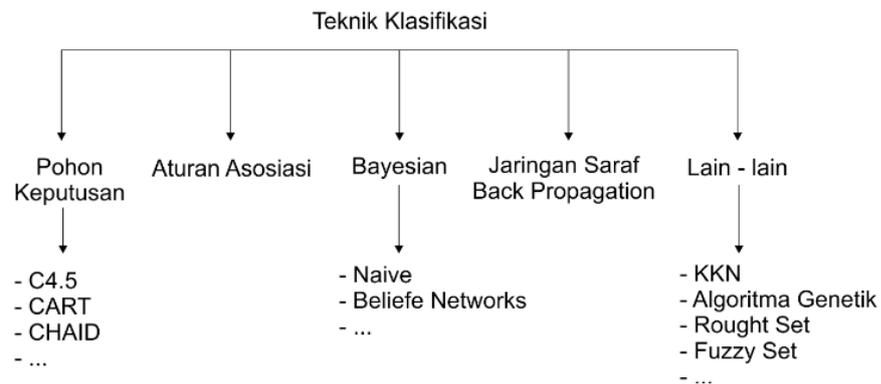
Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

#### 6. Presentasi Pengetahuan (Knowledge Presentation)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining (Han, 2006)

### **2.1.4 Klasifikasi**

Klasifikasi data adalah suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model Klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya. Teknik klasifikasi terbagi menjadi beberapa teknik yang diantaranya sebagai berikut : (Hamidah, 2013).



Gambar 2.1 Teknik Klasifikasi

## 2.2 Text Mining

*Text mining* (penambangan teks) adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda (Feldman & Sanger, 2007).

*Text mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, *clustering*, *information extraction* dan *information retrieval* (Berry & Kogan, 2010).

Jadi *text mining* dapat didefinisikan sebagai suatu proses menggali informasi dari sekumpulan dokumen yang bertujuan untuk mendapatkan informasi yang berguna dari sekumpulan dokumen.

## 2.3 Teknik Evaluasi Dalam Data Mining

Pada data mining untuk mengukur atau ada beberapa cara untuk mengukur kinerja dari model yang dihasilkan salah satunya menggunakan confusion matriks (akurasi) *Confusion matrix* adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Presisi atau *confidence* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* atau *sensitivity* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar.

Tabel 2.3 Model *Confusion Matrix*

Correct Classification	Classified as	
	+	-
+	True positives (A)	False negatives (B)
-	False positives (C)	True negatives (D)

Perhitungan akurasi dengan tabel *confusion matrix* adalah sebagai berikut:

$$\text{Akurasi} = (A+D)/(A+B+C+D)$$

Presisi didefinisikan sebagai rasio item relevan yang dipilih terhadap semua item yang terpilih. Presisi dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut. Rumus presisi adalah:

$$\text{Presisi} = A/(C+A)$$

*Recall* didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. *Recall* dihitung dengan rumus:

$$\text{Recall} = A/(A+D)$$

Presisi dan *Recall* dapat diberi nilai dalam bentuk angka dengan menggunakan perhitungan persentase (1-100%) atau dengan menggunakan bilangan antara 0-1. Sistem rekomendasi akan dianggap baik jika nilai presisi dan *recall*nya tinggi.

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positive* sebagai garis horizontal dan *true positive* sebagai garis vertikal. AUC (*the area under curve*) dihitung untuk mengukur perbedaan performansi metode yang digunakan. ROC memiliki tingkat nilai diagnosa yaitu:

- a. Akurasi bernilai 0,90 – 1,00 = *excellent classification*
- b. Akurasi bernilai 0,80 – 0,90 = *good classification*
- c. Akurasi bernilai 0,70 – 0,80 = *fair classification*
- d. Akurasi bernilai 0,60 – 0,70 = *poor classification*
- e. Akurasi bernilai 0,50 – 0,60 = *failure*

## **2.4 Sentiment Analysis**

*Sentiment analysis* adalah mengekstraksi pendapat, sentimen, evaluasi dan emosi orang tentang suatu topik tertentu yang tertulis menggunakan teknik pemrosesan bahasa alami. Sejumlah karya-karya besar lainnya menyebutkan *sentiment analysis* fokus pada aplikasi spesifik yang mengklasifikasikan mengenai opini positif, negatif dan netral (Sentiaji & Bachtiar, 2015).

*Sentiment analysis* atau yang disebut juga dengan *opini mining* merupakan analisis yang bertujuan untuk melihat opini masyarakat atau kelompok mengenai entitas tertentu (Apriliyanti P, 2015).

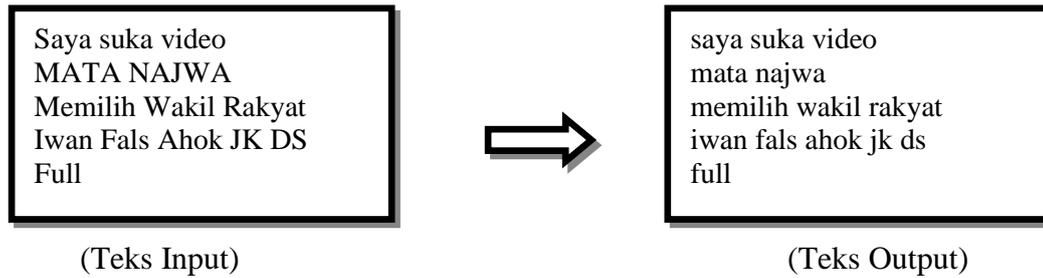
Jadi *sentiment analysis* atau *opinion mining* ini mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining*. Secara umum, bertujuan untuk menentukan *attitude* seorang penulis berkenaan dengan topik tertentu. *Attitude* mungkin penilaian atau evaluasi, pernyataan afektif (pernyataan emosional saat menulis) atau komunikasi emosional dimaksud (efek emosional penulis inginkan terhadap pembaca).

## **2.5 Preprocessing**

Tahap *preprocessing* diperlukan untuk membersihkan data dari teks yang tidak diperlukan, dimana data teks yang tidak terstruktur akan diubah menjadi data teks yang terstruktur atau semi terstruktur. Tahap dari *preprocessing* untuk mengolah data yaitu *case folding*, *convert emoticon*, *cleansing*, *tokenizing*, *stopword removal* dan *stemming*.

### **1. Case Folding**

*Case folding* merupakan tahapan merubah bentuk kata-kata menjadi sama bentuknya, baik semuanya menjadi *lower case* ataupun menjadi *upper case* (Sentiaji & Bachtiar, 2015). Berikut ini merupakan contoh *case folding* dapat dilihat pada gambar 2.2 :



Gambar 2.5.1 Contoh *Case Folding*

2. *Convert Emoticon*

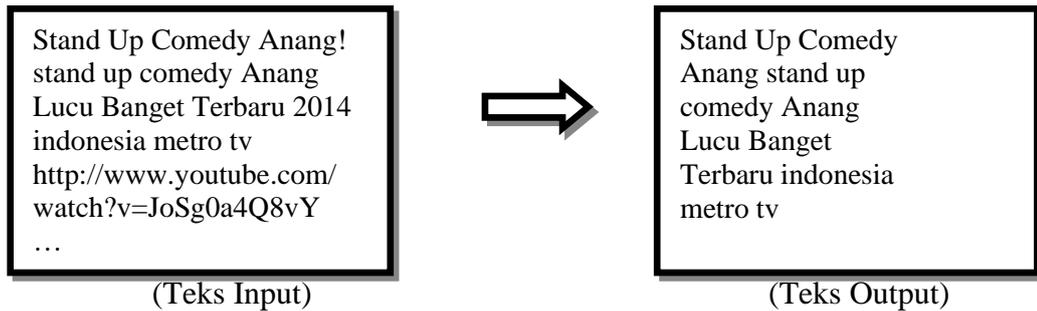
*Convert Emoticon* adalah proses untuk mengubah *emoticon* yang terdapat pada teks menjadi sebuah *string* (Jayanti. dkk, 2016). *Emoticon* merupakan salah satu cara pengungkapan ekspresi perasaan secara tekstual. Hal tersebut akan membantu menentukan sentimen dalam suatu kalimat atau teks (Sentiaji & Bachtiar, 2015). Berikut ini merupakan contoh *convert emoticon* dapat dilihat pada tabel 2.1 :

Tabel 2.5 Daftar *Convert Emoticon*

Emoticons	Deskripsi
:) :-) ;) ;-) ^_^ :P o.O :] :-] :3 :c) :> :[] 8) :^) :~) O:)	Senang
:v :D :-D B) B  8D x-D xD ==D =D ==3 B^D	Tawa
:-  :@ >:( 3:) :-?	Benci
>:O >:( -_- :-( :( :/ :'( ;( :-c :c :-< :~C :< :-[ :[	Kecewa
<3 :* (y)	Suka

3. *Cleansing*

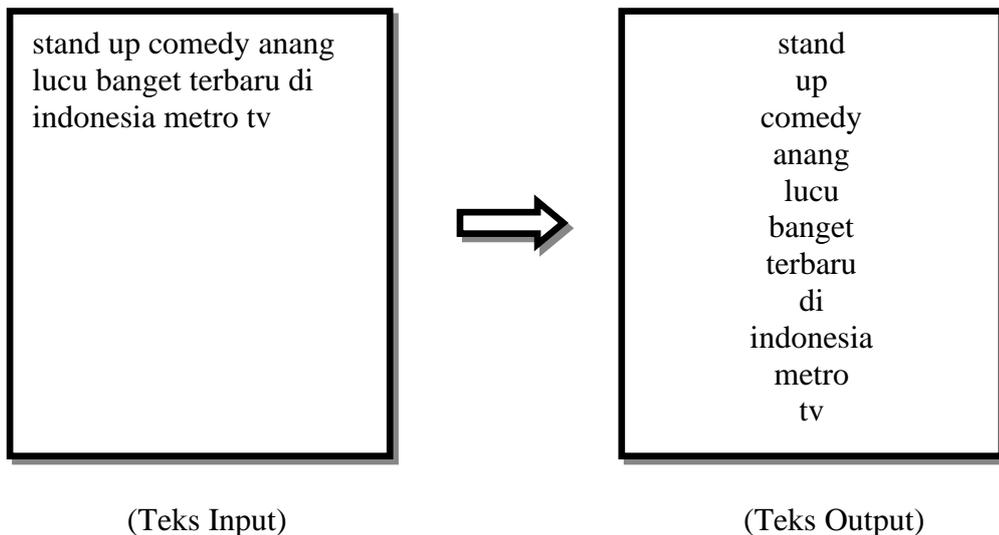
*Cleansing* merupakan proses membersihkan kata-kata yang tidak diperlukan untuk mengurangi noise (Sentiaji & Bachtiar, 2015). Kata yang dihilangkan adalah URL, *hashtag* (#), *username* (@username), email dan tanda baca (!@#%\$^&\*()\_+=.,<>?/|{}[]1234567890). Berikut ini merupakan contoh *cleansing* dapat dilihat pada gambar 2.3 :



Gambar 2.5.2 Contoh *Cleansing*

4. *Tokenizing*

*Tokenizing* berguna untuk memisahkan deretan kata menjadi beberapa urutan yang terpotong oleh spasi atau karakter spesial. (Sentiaji & Bachtiar, 2015). Berikut merupakan contoh *tokenizing* dapat dilihat pada gambar 2.4 :



Gambar 2.5.3 Contoh *Tokenizing*

5. *Stopword Removal*

*Stopword removal* berperan untuk membuang kata-kata yang sering muncul dan bersifat umum, kurang menunjukkan relevansinya dengan teks (Sentiaji & Bachtiar, 2015). Kata-kata yang akan dibuang tersebut didefinisikan dalam *stopword list*. Contoh beberapa kata yang sering masuk ke dalam *stopword list* adalah “sebuah”, “yang”, “itu”, “di” dan lain-lain. Berikut ini merupakan contoh proses *stopword removal* dapat dilihat pada gambar 2.5 :



## **2.6 Pengertian *Social Media***

Menurut Chris Brogan (2010:11) dalam bukunya yang berjudul *Social Media 101 Tactic and Tips to Develop Your Business Online* mendefinisikan *Social media* sebagai berikut: “*Social media is a new set of communication and collaboration tools that enable many types of interactions that were previously not available to the common person*”. (Sosial media adalah satu set baru komunikasi dan alat kolaborasi yang memungkinkan banyak jenis interaksi yang sebelumnya tidak tersedia untuk orang biasa).

*Social media* menurut Dailey (2009:3) adalah konten online yang dibuat menggunakan teknologi penerbitan yang sangat mudah diakses dan terukur. Paling penting dari teknologi ini adalah terjadinya pergeseran cara mengetahui orang, membaca dan berbagi berita, serta mencari informasi dan konten. Ada ratusan saluran *social media* yang beroperasi di seluruh dunia saat ini, dengan tiga besar facebook, LinkedIn, dan twitter.

### **2.6.1 Karakteristik *Social Media***

Menurut Hadi Purnama (2011:116) *social media* mempunyai beberapa karakteristik khusus diantaranya :

1. Jangkauan (*reach*): daya jangkauan *social media* dari skala kecil hingga khalayak global.
2. Aksesibilitas (*accessibility*): *social media* lebih mudah diakses oleh publik dengan biaya yang terjangkau.
3. Penggunaan (*usability*): *social media* relatif mudah digunakan karena tidak memerlukan keterampilan dan pelatihan khusus.
4. Aktualitas (*immediacy*): *social media* dapat memancing respon khalayak lebih cepat.
5. Tetap (*permanence*): *social media* dapat menggantikan komentar secara instan atau mudah melakukan proses pengeditan.

Keterkaitan dalam pembahasan ini peneliti menggunakan karakteristik *social media* karena pembahasan ini membahas mengenai *twitter* yang merupakan *social media*, dan bagaimana *twitter* mendukung kinerja *Public Relations* TRANS TV dalam melakukan publikasi terhadap program-program baru TRANS TV. Dan

bagaimana karakteristik ini mendukung pemakaian *twitter* oleh *followers*-nya saat ini.

### **2.6.2 Jenis Social Media**

Menurut Mayfield yang di kutip oleh Muhammad Badri (2011:133) menyebutkan saat ini ada tujuh jenis *social media*, namun inovasi dan perubahan terus terjadi. *Social media* yang ada saat ini :

1. Jejaring sosial seperti *facebook*, *myspace* dan *bebo*. Situs ini memungkinkan orang untuk membantu halaman *web* pribadi dan terhubung dengan teman-temannya untuk berbagi konten komunikasi.
2. *Blog*, merupakan bentuk terbaik dari media sosial, berupa jurnal *online* dengan pemuatan tulisan terbaik, yaitu tulisan terbaru ada di halaman terdepan.
3. *Wikis* seperti *Wikipedia* dan ensiklopedia *online website*. *Wikis* memperoleh siapa saja untuk mengisi atau mengedit informasi didalamnya, bertindak sebagai sebuah dokumen atau database komunal.
4. *Podcasts*, menyediakan *file-file audio* dan *video* dengan berlangganan melalui layanan seperti *Itunes* dari *Apple*.
5. Forum, area untuk diskusi *online*, seputar topik dan minat tertentu. Forum sudah ada sebelum media sosial dan menjadi komunitas *online* yang kuat dan populer.
6. Komunitas konten seperti *flickr* (untuk berbagi foto), *del.icio.us* (*link bookmarked*) dan *youtube* (*video*). Komunitas ini mengatur dan berbagi jenis konten tertentu.
7. *Microblogging*, situs jejaring sosial dikombinasikan *blog*, dimana sejumlah kecil konten (*update*) didistribusikan secara *online* dan melalui jaringan *mobile phone*, *twitter* adalah pemimpin layanan ini.

Dari definisi di atas peneliti menarik kesimpulan bahwa *social media* mempunyai jenis-jenis yang beraneka ragam dan mempunyai fungsi serta cara penggunaan yang berbeda-beda. Tetapi masih dengan tujuan yang sama untuk bertukar informasi dan mudah untuk diakses. Dalam penelitian ini, peneliti menggunakan *twitter* sebagai bahan penelitian.

### **2.6.3 Pengertian *Twitter***

Menurut Chris Brogan (2010:99) dalam bukunya yang berjudul *Social Media 101 Tactic and Tips to Develop Your Business Online* mendefinisikan *twitter* sebagai berikut:

*“Twitter is a great place to share ideas, an idea bank, a place to gather information, to inspire thoughts, or to see what your friends are doing”*. (Twitter adalah tempat yang tepat untuk berbagi ide, bank gagasan, tempat untuk mengumpulkan informasi, untuk menginspirasi pikiran, atau untuk melihat apa yang teman anda lakukan).

*Twitter* merupakan *mikroblog* paling populer di Indonesia. Mikroblog ini memungkinkan pengguna untuk mengirim dan membaca pesan yang disebut kicauan (*tweets*), berupa teks maksimal 140 karakter yang ditampilkan pada halaman profil pengguna. (Muhammad Badri, 2011:140)

### **2.6.4 Facebook**

*Facebook* merupakan situs media sosial yang memiliki pengguna aktif terbanyak di dunia. Sejak berdirinya Facebook pada tahun 2004 hingga Juni 2014, Facebook telah memiliki 1,32 Milyar pengguna aktif. Dimana Indonesia menduduki peringkat ke empat setelah Amerika, Brazil dan India, dengan jumlah 65 juta pengguna Facebook aktif. Sebanyak 33 juta pengguna aktif per harinya, 55 juta pengguna aktif yang memakai perangkat *mobile* dalam pengaksesannya per bulan dan sekitar 28 juta pengguna aktif yang memakai perangkat *mobile* per harinya (kominfo.go.id).

## **2.7 Algoritma Naive Bayes Classifier**

Pendekatan Naive Bayes merupakan sebuah metode klasifikasi yang mengacu pada teorema Bayes. Teorema Bayes digunakan untuk menghitung probabilitas ketidakpastian data (Russel & Norvig 2010). Proses pendekatan Naive Bayes Classifier mengasumsikan bahwa ada atau tidaknya suatu fitur pada suatu kelas tidak berhubungan dengan ada atau tidaknya fitur lain di kelas yang sama (Setiawan & Ratnasari 2015). Pada saat klasifikasi, pendekatan Bayes akan

menghasilkan tabel kategori yang paling tinggi nilai probabilitasnya yaitu VMAP (Maximum Apriori Probability) dengan atribut inputan  $G_1, G_2, \dots, G_n$  (Aribowo 2010).

Persamaan dari teorema Bayes adalah (Bustami, 2013) :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Dimana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$  : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$  : Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$  : Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$  : Probabilitas X

Untuk menjelaskan metode Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode Naive Bayes di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)}$$

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

$$\textit{Posterior} = \frac{\textit{prior} \times \textit{likelihood}}{\textit{evidence}}$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan  $(C|F_1, \dots, F_n)$  menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned}
 P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\
 &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\
 &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\
 &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\
 &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1})
 \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing petunjuk  $(F_1, F_2, \dots, F_n)$  saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk  $i \neq j$ , sehingga

$$P(F_i|C, F_j) = P(F_i|C)$$

Persamaan di atas merupakan model dari teorema Naive Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss :

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Di mana :

P : Peluang

Xi : Atribut ke i

$x_i$  : Nilai atribut ke  $i$

$Y$  : Kelas yang dicari

$y_i$  : Sub kelas  $Y$  yang dicari

$\mu$  : mean, menyatakan rata – rata dari seluruh atribut

$\sigma$  :Deviasi standar, menyatakan varian dari seluruh atribut.

## **BAB III METODE PENELITIAN**

### **3.1 Metode Penelitian**

Metode penelitian yang digunakan pada penelitian ini sebagai berikut.

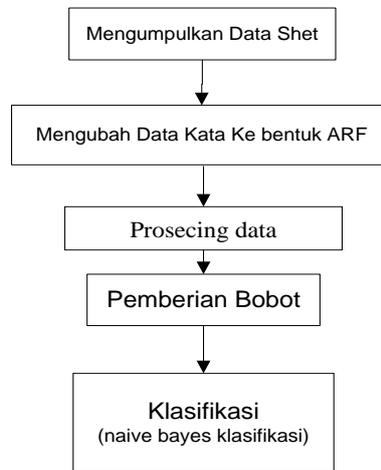
#### **3.1.1 Metode Pengumpulan Data**

Penelitian ini melakukan metode pengumpulan data sehingga dapat melengkapi data yang ada. Adapun tahapan metode yang digunakan adalah sebagai berikut :

- a. Wawancara : melakukan wawancara langsung dengan pihak terkait untuk mendapatkan keterangan-keterangan yang diperlukan sebagai bahan penulisan laporan.
- b. Observasi : mendapatkan data-data dan fakta dari pengamatan langsung di grup Facebook dan twiter fauzan sibron, bustami zainudin, ridho fikardo dan mahfud md.
- c. Studi Pustaka : dilakukan dengan cara membaca, mengutip dan membuat catatan yang bersumber pada bahan-bahan pustaka yang mendukung dan berkaitan dengan penelitian ini khususnya dalam pengembangan Sistem Pendukung Keputusan. Selanjutnya dengan cara mempelajari dan memahami jurnal dan buku-buku referensi, yang berhubungan dengan masalah yang akan dibahas dalam karya ilmiah ini. Hal ini dimaksudkan agar penulis memiliki landasan teori yang kuat.

### 3.2. implementasi

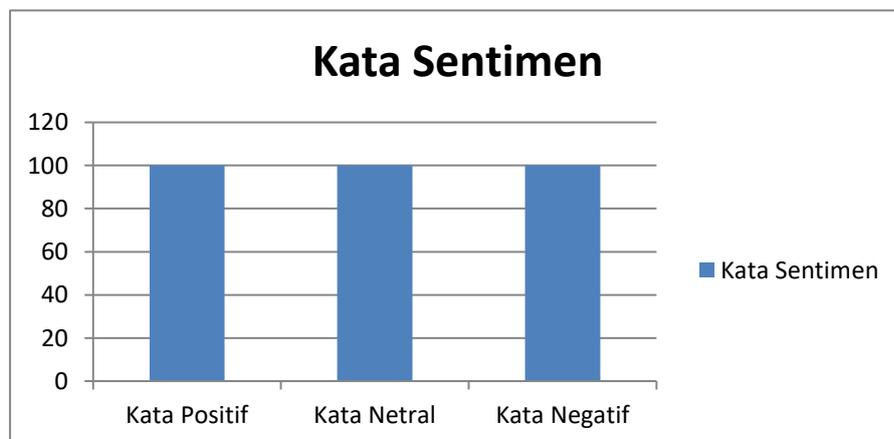
Implementasi penelitian ini di lakukan dengan langkah seperti gambar 3.2



Gambar 3.2 Diagram Alir Kerja Penelitian

Proses awal pada tahap implementasi adalah pengumpulan data set yang akan digunakan baik untuk testing maupun training. Pengumpulan data set dengan cara memanfaatkan komentar dari facebook dan twiter rentang waktu 6 bulan melalui teknik manual data kemudian dilakukana preprocessing data untuk menjadikan sederhana dimensi dari data set

Dari set yang di ambil terdiri dari 100 kata positif, 100 kata netral dan 100 kata negatif.



3.1 Tabel Data Shet Sentimen

### 3.1.3 Metode Pengembangan Perangkat Lunak

#### 3.1.3.1 Perencanaan

Pada tahapan ini, penulis mengumpulkan semua kebutuhan elemen sistem kemudian dialokasikan pada sistem yang ada, dan pada tahap ini berkaitan dengan penentuan kebutuhan pengguna dan perencanaan proyek. Perencanaan yang harus dipersiapkan adalah melakukan pengumpulan data yang berkenaan dengan analisis sentimen pada media sosial grup Facebook Muhammad Ridho Ficardo, mahfud md, fauzan sibron dan bustami zainudin.

#### 3.1.3.2 Analisis

a. Analisis Kebutuhan *Hardware*

Spesifikasi *hardware* yang digunakan untuk membuat suatu aplikasi media pembelajaran Aquascape, sebagai berikut:

1. *ProcessorCore i3*2.53 GHz dan 2.53 GHz
2. Monitor 14".
3. Ram 2Gb.
4. VGA 1Gb
5. *Harddisk* 20 Gb.
6. *Keyboard* dan *mouse* Optik.

b. Analisis Kebutuhan *Software*

Untuk menjalankan program ini juga dibutuhkan spesifikasi *software* yang digunakan sebagai pendukung sistem. Adapun Spesifikasi *Software* yang digunakan antara lain :

1. Perangkat lunak sistem operasi yang digunakan adalah *Microsoft Windows-7*.
2. Perangkat lunak aplikasi yang digunakan
  - a. Microsoft Office Excel Professional Plus 2007 Software ini digunakan sebagai media penulisan datasheet.
  - b. Notepad++ adalah suatu text editor yang berjalan pada Operating System(OS) Windows. Notepad++ disini menggunakan komponen-

komponen *Scintilla* agar dapat menampilkan dan menyunting text dan berkas source code berbagai bahasa pemrograman.

- c. Weka adalah aplikasi data mining open source berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi / formulasi dari sekumpulan data sampling. Walaupun kekuatan Weka terletak pada algoritma yang makin lengkap dan canggih, kesuksesan data mining tetap terletak pada faktor pengetahuan manusia implementornya. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan.

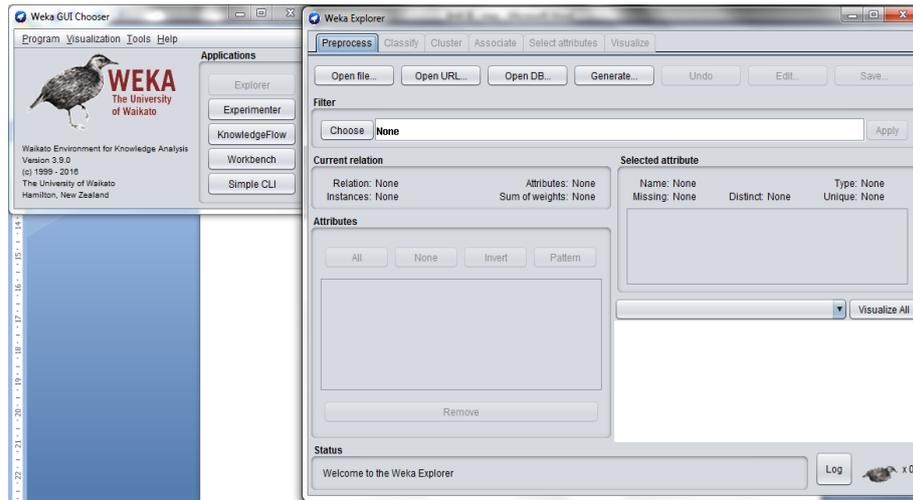


**3.2 Gambar Aplikasi WEKA**

### **3.1.1.1 Pembuatan**

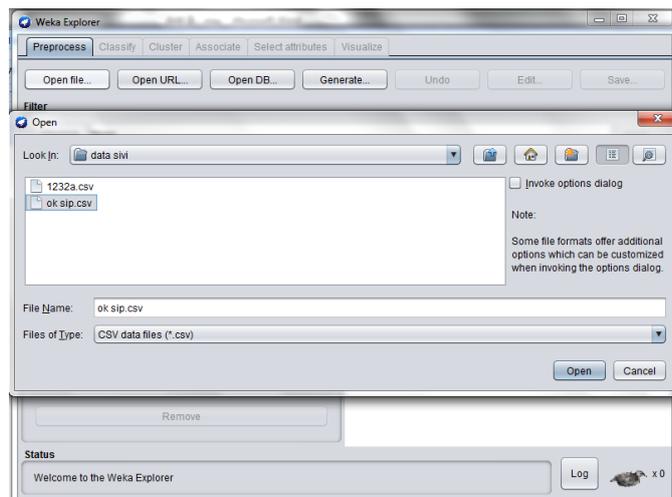
#### **1. Mengubah File CSV menjadi ARF**

Dalam proses ini, data yang sudah di buat dalam excel dengan format *CSV* diubah menjadi format *ARF* dengan menggunakan tools Weka 3.90. Dapat di liat pada gaambar berikut:



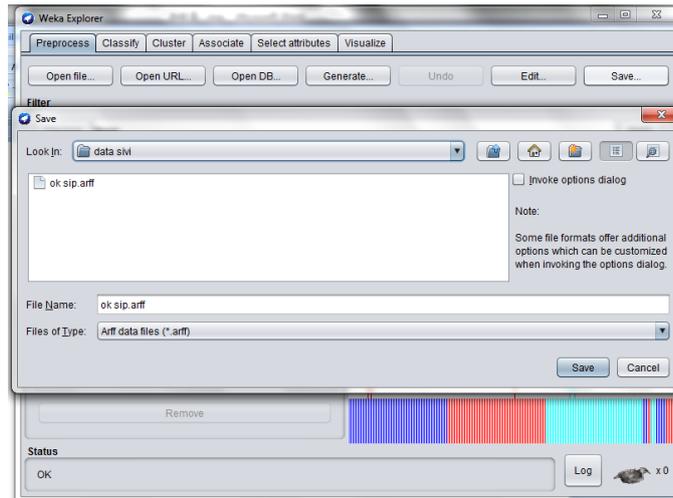
**3.3 Gambar Open Exploler pada Aplikasi WEKA**

setelah open exploler pada weka, kita membuka open file untuk mencari data shet dalam bentuk excel agar dapat di ubah dalam bentuk ARF.



**3.4 Gambar Open File pada Aplikasi WEKA**

data pada excel akan tampil pada aplikasi weka yang di gambarkan dalm bentuk warna, sehingga dapat membagi setiap data shet yang ada. Setelah data shet tampil barulah kita mengubah file CSV pada excel dengan mengubah format ke ARF menggunakan weka.



**3.5 Gambar Mengubah format datshet ke ARF**

2. *Pre-processing* (Persiapan data)

Melakukan *preprocessing* data kata sentimen. *Preprocessing* meliputi lower case tokens, normalization, tokenization *cleansing*, dan *filtering*. Semua tahapan preprocessing data dilakukan menggunakan tools WEKA 3.9.0. Tahapan-tahapan prepeocsing yang dilakukansebagai berikut:

a. *Lower Case Tokens*

berfungsi untuk membuat data tweet menjadi huruf kecil semua, misalnya dari huruf capital menjadi huruf kecil biasa.

b. *Normalization*

dilakukan untuk menormalkan kata-kata yang tidak baku, misalnya kata yang disingkat, kata gaul atau alay ke kata yang bakunya di Bahasa Indonesia.

c. *Cleansing*

yaitu proses menghapus simbolsymbol yang kurang penting dalam data

d. *kata sentimen*

yang bisa mengganggu proses klasifikasi nantinya. Proses ini dilakukan dengan menggunakan menu *delimiters* pada WEKA.

e. *Filtering*

dilakukan untuk menghapus katakata yang kurang penting atau kurang

berpengaruh terhadap proses klasifikasi nantinya. Proses ini dilakukan dengan menggunakan *stopword list*. *Stopword list* yang digunakan dalam penelitian ini adalah *stopword list* WEKA dan *stopword list*.

f. Hitung frekuensi kemunculan kata

Hal ini dimaksudkan untuk menghitung berapa banyak kata yang sama di dalam kata sentimen yang dapat menggambarkan ciri dari masing-masing kata sentimen.

g. Hitung Probabilitas kata

Menghitung probabilitas kata dilakukan untuk mendapatkan *term* dengan nilai yang lebih penting dan dianggap relevan untuk dijadikan kata kunci. Proses pembobotan menggunakan algoritma *Naive Bayes*, dalam proses perhitungan penulis mengasumsikan bahwa:

a.  $W_k$  (kata),  $V_j$  (kategori dokumen). Jadi  $(P(W_k|V_j))$  sama dengan  $n_k$  (frekuensi munculnya  $W_k$ ) ditambah 1 kemudian dibagi dengan  $n$  (banyaknya seluruh kata dalam kategori) dan ditambah dengan banyaknya kosakata  $|kosakata|$ . Diperoleh persamaan sebagai berikut:

$$P(W_k|V_j) = \frac{n_k+1}{n+|Kosakata|}$$

**Contoh :**

Diketahui : nPositif: 100  
nNetral: 100  
nNegatif: 100

Kata Positif

$$P(lanjut|Positif) = \frac{9+1}{100+300} = \frac{10}{400} = 0,04$$

$$P(lanjut|netral) = \frac{0+1}{100+300} = \frac{1}{400} = 0,0025$$

$$P(lanjut|negatif) = \frac{0+1}{100+300} = \frac{1}{400} = 0,0025$$

kata netral

$$P(Gubernur|Positif) = \frac{0+1}{100+300} = \frac{1}{400} = 0,0025$$

$$P(\text{Gubernur}|\text{Netral}) = \frac{34+1}{100+300} = \frac{35}{400} = 0.0875$$

$$P(\text{Gubernur}|\text{Negatif}) = \frac{0+1}{100+300} = \frac{1}{400} = 0,0025$$

Kata negatife

$$P(\text{Desa}|\text{Positif}) = \frac{0 + 1}{100 + 300} = \frac{1}{400} = 0,0025$$

$$P(\text{Desa}|\text{Netral}) = \frac{0+1}{100+300} = \frac{1}{400} = 0,0025$$

$$P(\text{Desa}|\text{Negatife}) = \frac{26+1}{100+300} = \frac{27}{400} = 0.0675$$

**Tabel 3.2 Contoh Daftar Kata Sentimen**

No	Kata	Positif	Netral	Negatif	Probab Positif	Probab Netral	Probab negatif
1	Lanjut	9	0	0	0,4	0,0025	0,0025
2	Gubernur	0	34	0	0,0025	0,0875	0,0025
3	Desa	0	0	26	0,0025	0,0025	0,0675
4	kebebasan	24	0	0	0.0625	0,0025	0,0025
5	kerukunan	35	0	0	0.09	0,0025	0,0025
6	kegembiraan	20	0	0	0.0525	0,0025	0,0025
7	indah	26	0	0	0.0675	0,0025	0,0025
8	asri	43	0	0	0.11	0,0025	0,0025
9	elok	0	23	0	0,0025	0.06	0,0025
10	kami	0	22	0	0,0025	0.0575	0,0025
11	orang	0	43	0	0,0025	0.11	0,0025
12	tanjung	0	16	0	0,0025	0.0425	0,0025
13	kecamatan	0	20	0	0,0025	0.0525	0,0025
14	lampung	0	46	0	0,0025	0.1175	0,0025
15	bukan	0	0	44	0,0025	0,0025	0.1125
16	hangus	0	0	52	0,0025	0,0025	0.1325
17	kasus	0	0	23	0,0025	0,0025	0.06
18	merendahkan	0	0	27	0,0025	0,0025	0.07
19	kalah	0	0	31	0,0025	0,0025	0.08
20	bakar	0	0	15	0,0025	0,0025	0.04

- b. Setelah mendapatkan hasil dari probabilitas kata, kemudian akan menghitung probabilitas dari dokumen sampel atau contoh, diasumsikan  $P(V_j)$  (probabilitas kategori dokumen) sama dengan  $\frac{docs_j}{j}$  (jumlah dokumen

setiap kategori) dibagi dengan  $|contoh|$  (jumlah dokumen yang digunakan sebagai data *training* dari seluruh kategori). Diperoleh persamaan berikut:

$$P(V_j) = \frac{docs_j}{|contoh|}$$

**Contoh** : perhitungan probabilitas kata, dengan asumsi terdapat 3 kata sentimen.

1 dokumen kata positif, 1 dokumen kata netral dan 1 dokumen kata negatif :

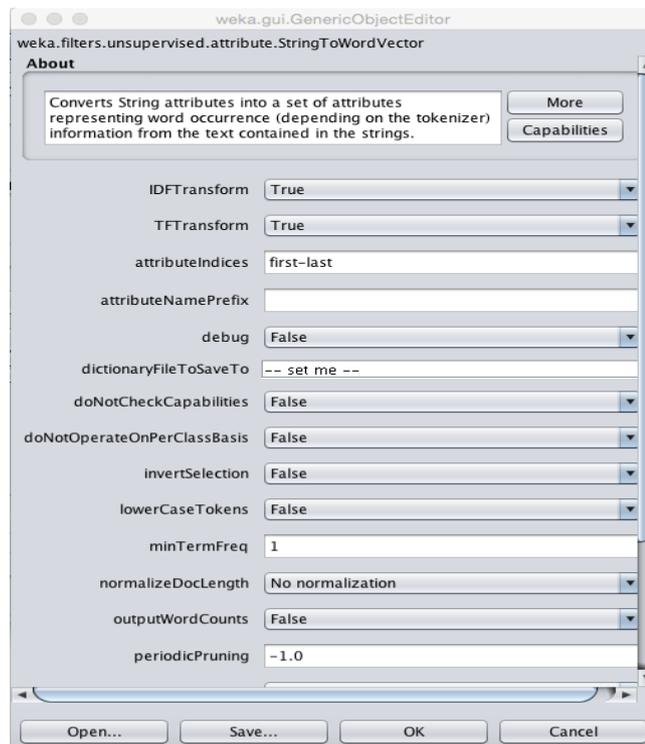
$$P(Kriminal) = \frac{1}{3} = 0,33$$

$$P(Olahraga) = \frac{1}{3} = 0,33$$

$$P(Bisnis) = \frac{1}{3} = 0,33$$

### 3. Pemberian Bobot

Pada tahap selanjutnya adalah memberikan bobot pada tiap-tiap kata (*term*). Pembobotan dilakukan untuk mendapatkan nilai dari kata yang berhasil diekstrak. Metode yang digunakan untuk pemberian bobot dalam penelitian ini adalah TFIDF (*Term Frequency – Inverse Document Frequency*). Karena metode ini paling cocok jika dikombinasi dengan metode klasifikasi *Naïve Bayes Classifier* (NBC).



**Gambar 3.7.** TF-IDF (*Term Frequency – Inverse Document Frequency*)

4.. *Klasifikasi*

Pada penelitian ini, tahap klasifikasi data menggunakan bantuan tools WEKA 3.9.0. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naïve Bayes Classifier (NBC)*. *Naïve Bayes Classifier (NBC)* adalah metode *classifier* berdasarkan probabilitas dan *Teorema Bayesian* dengan asumsi bahwa setiap variabel X bersifat bebas (*independence*). Dengan kata lain, *Naïve Bayesian Classifier (NBC)* mengasumsikan bahwa keberadaan sebuah atribut (variabel) tidak ada kaitannya dengan beradaan atribut (variabel) yang lain. Berikut adalah Rumusnya.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Pada proses klasifikasi data diuji dengan menggunakan metode 10 fold cross validation. Jadi dataset akan dibagi menjadi dua, yaitu 10 bagian dengan 9/10 bagian digunakan untuk proses training dan 1/10 bagian digunakan untuk proses testing. Iterasi berlangsung 10 kali dengan variasi data training dan testing menggunakan kombinasi 10 bagian data.

Pengujian	Datashet									
1	■									
2		■								
3			■							
4				■						
5					■					
6						■				
7							■			
8								■		
9									■	
10										■

**Gambar 3.8.** Ilustrasi 10 fold cross validation

### 3.3 . Evaluasi Hasil

Melakukan evaluasi performa *TP rate*, *FP rate*, *Precision*, *Recall* dan *F-measure* dari eksperimen yang telah dilakukan. Evaluasi dilakukan dengan menggunakan *Confusion Matrix* yaitu *true positive rate (TP rate)*, *true negative rate (TN rate)*, *false positive rate (FP rate)* dan *false negative rate (FN rate)* sebagai indikator. *TP rate* adalah persentase dari kelas positif yang berhasil diklasifikasi sebagai kelas positif, sedangkan *TN rate* adalah persentase dari kelas negatif yang berhasil diklasifikasi sebagai kelas negatif. *FP rate* adalah kelas negatif yang diklasifikasi sebagai kelas positif. *FN rate* adalah kelas positif yang diklasifikasi sebagai kelas negatif.

**Table 3.3** Confusion Matrik

		Predicted		
		Positif	Negatif	Netral
Actual	Positif	18	20	62
	Negatife	31	11	58
	Netral	25	16	59

## **BAB IV**

### **HASIL PENELITIAN DAN PEMBAHASAN**

#### **4.1 Implementasi**

Implementasi merupakan suatu proses pelaksanaan atau penerapan dari sebuah rancangan yang telah disusun secara terperinci. Berikut ini merupakan mekanisme implementasi dari penelitian yang telah dilakukan:

##### **4.1.1 Skenario Pengujian**

Hasil berupa data penelitian yang telah diolah dan dituangkan dalam bentuk tabel, grafik, foto atau gambar. Pembahasan berisi hasil analisis dan hasil penelitian yang dikaitkan dengan struktur pengetahuan yang telah mapan (tinjauan pustaka yang diacu oleh penulis), dan memunculkan teori– teori baru atau modifikasi terhadap teori – teori yang telah ada.

Dataset pada penelitian ini menggunakan format ARFF yang dikumpulkan dari Twitter dan facebook dengan metode manual melihat dari setiap komentar dari media sosial Twitter dan facebook. Data yang diambil hanya tweet dalam bahasa Indonesia, yaitu pada halaman fauzan sibron, bustami zainudin, ridho fikardo, dan mahmud amien. Data diambil secara acak baik dari user biasa ataupun media online. Dataset yang digunakan sebanyak 300 kata, data dibagi secara seimbang (balanced) setiap kelasnya, karena dengan data yang tidak seimbang (imbalanced), klasifikasi yang dibangun memiliki kecenderungan untuk mengabaikan minority class. Data dibagi menjadi kata positif 100 kata, kata negatif 100 kata dan kata netral 100 kata.

Untuk mengetahui akurasi, Analisis Sentimen Pada Media Sosial diklasifikasi menggunakan metode *Naïve Bayes Classifier (NBC)* dengan perangkat lunak WEKA versi 3.9.0. WEKA menggunakan tipe dokumen *Atribut-Relation File Format (ARFF)* sebagai masukan untuk melakukan klasifikasi data.

Sekarang mengetahui uji training, data di masukan ke weka 3.9.0 dan dilihat hasilnya. Hasil lengkap dapat di lihat pada tabel berikut;

**Tabel 4.1 Jumlah Data**

Jumlah			
Mean	32.6088	32.0155	31.162
Std.dev	12.0915	12.0079	11.1565
Weight sum	100	100	100
Precision	1.0408	1.0408	1.0408

**Tabel 4.2 Summary**

Correctly Classified Instances	297	99%
Incorrectly Classified Instances	3	1%
Kappa statistic	0.985	
Mean absolute error	0.3319	
Root mean squared error	0.3532	
Relative absolute error	74.6844 %	
Root relative squared error	74.9162 %	
Total Number of Instances	300	

**Tabel 4.3 Detailed Accuracy By Class**

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.980	0.005	0.990	0.980	0.985	0.977	1.000	1.000	Positif
0.990	0.010	0.980	0.990	0.985	0.978	1.000	1.000	Negatif
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Netral
0.990	0.005	0.990	0.990	0.990	0.985	1.000	1.000	Weighted Avg

**Tabel 4.4 Confusion Matrix**

classified as	Positif	Netral	Negatif
Positif	98	2	0
Netral	1	99	0
Negatif	0	0	100

Dari beberapa tabel dia atas, dengan menggunakan metode *naive bayes classification* untuk mengecek data training hasil yang dapat sangat memuaskan dengan tingkat akurasi 99% atau 297 kata yang benar dan hanya 3% kesalahan yaitu sebanyak 3 kata.

Setelah mendapatkan data uji, kita akan mengujinya menggunakan data testing. Data testing terdiri dari 10 % dari jumlah data atau sebanyak 30 kata. hasil dapat di lihat pada tabel berikut:

**Tabel 4.5 Jumlah Data**

Jumlah			
Mean	32.6088	32.0155	31.162
Std.dev	12.0915	12.0079	11.1565
Weight sum	100	100	100
Precision	1.0408	1.0408	1.0408

**Tabel 4.6 Summary**

Correctly Classified Instances	30	100%
Incorrectly Classified Instances	0	0%
Kappa statistic	1	
Mean absolute error	0.3408	
Root mean squared error	0.3629	
Relative absolute error	76.6862 %	
Root relative squared error	76.9876 %	
Total Number of Instances	30	

**Tabel 4.7 Detailed Accuracy By Class**

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Positif
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Negatif
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Netral
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Weighted Avg

**Tabel 4.8 Confusion Matrix**

classified as	Positif	Netral	Negatif
Positif	10	0	0
Netral	0	10	0
Negatif	0	0	10

Dari beberapa tabel dia atas, dengan menggunakan metode *naive bayes classification* untuk mengecek data sethasil yang dapat sangat memuaskan dengan tingkat akurasi 100 atau 10 dari 10 kata yang benar dan 0% kesalahan yaitu sebanyak 0 kata atau bia di sebut juga tidak ada kesalahan dan pendeteksiannya.

Hasil dari proses klasifikasi kemudian diuji dengan menggunakan metode *10-fold crossvalidation*, data dibagi mejadi 10 bagian dengan 9/10 bagian digunakan untuk proses *training* dan 1/10 bagian digunakan untuk proses *testing*.

Iterasi berlangsung 10 kali dengan variasi data *training* dan *testing* menggunakan kombinasi 10 bagian data. Perbandingan hasil dari metode klasifikasi *Naïve Bayes Classifier (NBC)* dengan kata sentimen yang sudah di tentukan sebelumnya. Selengkapnya bsia di lihat pada tabel berikut:

**Tabel 4.9 Jumlah Data**

Jumlah			
Mean	32.6088	32.0155	31.162
Std.dev	12.0915	12.0079	11.1565
Weight sum	100	100	100
Precision	1.0408	1.0408	1.0408

**Tabel 4.10 Summary**

Correctly Classified Instances	88	29.3333%
Incorrectly Classified Instances	212	70.6667%
Kappa statistic	-0.06	
Mean absolute error	0.4454	
Root mean squared error	0.4736	
Relative absolute error	100.2188 %	
Root relative squared error	100.4623 %	
Total Number of Instances	300	

**Tabel 4.11 Detailed Accuracy By Class**

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.180	0.280	0.243	0.180	0.207	-0.109	0.463	0.309	Positif
0.110	0.110	0.234	0.110	0.150	-0.091	0.399	0.274	Negatif
0.590	0.590	0.330	0.590	0.423	-0.010	0.506	0.364	Netral
0.293	0.353	0.269	0.293	0.260	-0.070	0.456	0.316	Weighted Avg

**Tabel 4.12 Confusion Matrix**

classified as	Positif	Netral	Negatif
Positif	18	20	62
Netral	31	11	58
Negatif	25	16	59

Tabel di atas berisi informasi mengenai nilai akurasi, presisi, recall, TP rate dan TN rate dari masing-masing uji coba yang telah dilakukan.

Bagian kolom berisi informasi analisis kata sentimen di sosial media. Sedangkan bagian baris berisi nilai akurasi, presisi, recall, TP rate dan TN rate dari masing-masing uji coba yang telah dilakukan. Dari proses data preprocessing menghasilkan sejumlah token yang kemudian digunakan sebagai input sebuah proses klasifikasi. Proses klasifikasi dilakukan menggunakan metode Naïve Bayes Classifier (NBC). Dari proses klasifikasi diperoleh nilai akurasi, presisi, recall, TP rate dan TN rate dari masing-masing uji coba.

## **4.2 Analisis Sistem**

Berdasarkan hasil pengujian terhadap sistem maka dapat dilakukan beberapa analisis. metode Naïve Bayes Classifier (NBC). Nilai akurasi dataset training mencapai 99%, nilai presisi 74.6844 %, nilai recall 77%, nilai TP rate 98,6% dan nilai TN rate 22,2%. Untuk dataset testing akurasinya mencapai 100%, nilai presisi 74,4%, nilai recall 76%, nilai TP rate 93,8% dan nilai TN rate 52,9%. Dataset testing mendapatkan akurasi paling tinggi karena dari 30 data yang terdiri dari 10 kata positif, 10 kata netral dan 10 kata negatif, 30 data berhasil diklasifikasi metode Naïve Bayes Classifier (NBC) dengan benar sesuai sentimennya yaitu sentimen positif.

Sedangkan untuk dataset Training, dari 300 data yang di uji, 297 data berhasil diklasifikasi metode Naïve Bayes Classifier (NBC) dengan benar sesuai sentimennya yaitu positif, negatif dan netral. Metode ini cenderung Naïve Bayes Classifier (NBC) lebih stabil karena berbasis probabilitas kemunculan kata dalam sebuah kalimat. Nilai akurasi adalah salah satu parameter penilaian dari metode yang telah digunakan, nilai akurasi didapat dari jumlah banyaknya data yang berhasil diklasifikasi dengan benar sesuai kelas sentimennya dari seluruh jumlah data yang diklasifikasi.

Nilai akurasi yang tinggi didapat ketika banyak data yang berhasil diklasifikasi dengan benar sesuai kelas sentimennya. Dapat diketahui juga nilai Precision dan Recall. Nilai Precision mengikuti nilai akurasi, nilai akurasi semakin tinggi maka akan diikuti nilai Precision yang tinggi juga, begitu

sebaliknya. Nilai Precision adalah jumlah data positif yang benar diklasifikasi sebagai data positif dibagi total data yang diklasifikasi sebagai data positif.

Sedangkan Nilai recall adalah jumlah data positif yang benar diklasifikasi sebagai data positif dibagi jumlah data positif sebenarnya. Kita juga dapat mengetahui nilai TP Rate dan TN Rate. TP Rate adalah nilai data positif yang diklasifikasi dengan benar sesuai kelas sentimennya, yaitu positif. Nilai TN Rate adalah nilai data sentimen yang diklasifikasi dengan benar sesuai kelas sentimennya, yaitu negatif.

Dari penelitian yang telah dilakukan, terbukti metode klasifikasi Naïve Bayes Classifier (NBC) dapat digunakan untuk mengklasifikasi sentimen sentimen (positif, netral dan negatif) tweet Bahasa Indonesia. Selanjutnya, dataset testing mendapat akurasi lebih tinggi dibandingkan dengan akurasi dataset training, dengan akurasi 100% dibanding 99%.

Dalam dataset testing, sentimen positif, negatife dan netralnya berhasil teridentifikasi semua dengan metode *naive bayes classification*. Sedangkan didataset training ada 297 kata yang teridentifikasi dan 3 kata lagi tidak. Jadi bisa disimpulkan, bahwan metode *naive bayes classification* dapat memberikan hasil yang cukup memuaskan atau cukup valid dalam menentukan analisis sentimen.

## **BAB V**

### **SIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan analisis yang dilakukan pada penelitian ini, maka dapat diambil kesimpulan yaitu:

4. Penelitian ini mengklasifikasikan sentimen pada opini Facebook dengan proses *Text Mining* menggunakan metode *Naive Bayes Classifier* (NBC). Tahapan klasifikasi sentimen yaitu tahap *training dan testing*. Pada tahap *training* terdapat proses seperti *pre-processing*, perhitungan probabilitas dan pengelompokan dengan algoritma *Naive Bayes Classifier* untuk menghasilkan kata kunci setiap kategori teks dan nilai probabilitasnya kemudian di *testing*.
5. Analisis hasil perhitungan metode *Naive Bayes Classifier* (NBC) dilakukan berdasarkan sumber data didapat dari grup Facebook fauzan sibron, bustami zainudin dan ridho fikardo dengan 300 data, 90% (270 kata) digunakan sebagai data training dan 10% (30 kata) sebagai data testing.
6. Hasil akurasi metode *Naive Bayes Classifier* (NBC) berdasarkan evaluasi hasil pengujian dengan dokumen acak menghasilkan probabilitas tertinggi sesuai dengan kata kunci dokumen pembelajaran dan menghasilkan presentase pengujian 100% untuk kategori positif, 100% kategori negatif dan 100% kategori netral.

#### **5.2 Saran**

Berdasarkan kesimpulan yang telah diuraikan, saran yang diberikan adalah untuk mengembangkan sistem ini lebih lanjut, dapat menambahkan proses pengujian secara langsung melalui aplikasi.

## DAFTAR PUSTAKA

- HertantoWidododkk. 1999. PAS (PanduanAkuntansiSyariah) PanduanPraktisOperasionalBaitulMaalWatTamwil (BMT). Bandung: Mizan.
- Ibrahim Lubis. 1995.EkonomiIslam SuatuPengantarJilid 2.Jakarta: KalamMulia.
- Jazuli,danYadiJanwari.2002. LembagalembagaPerekonomianUmat. Jakarta:PT. Raja GarfindoPersada.
- Jefri.2013. ImplementasiAlgoritma C4.5 DalamAplikasiUntukMemprediksiJumlahMahasiswa Yang Mengulang Mata Kuliah Di STMIK AMIKOM.Yogyakarta : Yogyakarta.
- Jiawei Han, MichelineKamber. 2001.*Data Mining Concept and Techniques*.
- Kasmir. 2001.ManajemenPerbankan. Jakarta : PT Raja GrafindoPersada.
- Kusrini, 2008.Computing For Humanity Algoritma C4.5.Yogyakarta.
- Kusrini, danEmhaTaufikLuthfi.2009.Algoritma Data Mining.Yogyakarta: Andi Offset.
- Muhammad.2002.Lembaga-lembagaKeuanganUmatKontemporer. Yogyakarta: UII Press.
- M.DawanRaharjo. 1999. Islam danTransformasiSosial-Ekonomi. Yogyakarta: PustakaPelajar.
- Muhammad Ridwan. 2004.ManajemenBaitulMaalWaTamwil (BMT). Yogyakarta: UII Press.
- MunawarohHolisatuldkk. 2013PerbandinganAlgoritma ID3 DanC5.0 DalamIdentifikasiPenjurusanSiswa SMA”, JurnalSarjanaTeknikInformatika, Vol. 1, No. 1, pp. 1-12, Juni.
- Prasetyo, Eko. 2012. Data Mining :KonsepdanAplikasiMenggunakanMatlab. Yogyakarta: CV. Andi Offset.
- Syafi’i Antonio. 2001. Bank Syari’ah Dari TeoriKePraktek. karta: PenerbitGemaInsani.

# LAMPIRAN

Kata Positif	Jumlah	Kata Negatif	Jumlah	Kata Netral	Jumlah
Lanjut	9	Desa	26	Gubernur	34
Satuperiodelagi	11	Kalau	44	Kami	22
Berbakhti	44	Gakguna	28	Orang	43
Setuju	33	Payah	30	Tanjung	16
Agung	23	Musibah	33	Kecamatan	20
Lagi	19	Rusuh	14	Lampung	46
Dukung	21	Penjarahan	10	Katakan	44
Semangat	31	Berdosa	33	Jadi	33
Salam	24	Emosi	36	Ketua	31
Juara	21	Luka	39	Dunia	14
Mohon	43	Jahat	20	Masyarakat	30
Semua	24	Belum	54	Acara	27
Terimakasih	38	Minta	22	Buka	19
Penuh	26	Sulit	19	Anggota	22
Lahir	45	Buang	16	Ditinggalkan	14
Jalan	40	Hilang	32	Contoh	46
Datang	12	Maling	38	Korban	25
Optimis	20	Pecah	21	Banyak	41
Yakin	32	Rusak	27	Waktu	17
Baik	40	Hoax	36	Pulang	21
Kembali	38	Korupsi	39	Mumpung	15
Cepat	21	Bingung	29	Jauh	38
Doa	15	Buang	21	Jatuh	29
Tolong	44	Kemarahan	32	Jangan	37
Senang	29	Jarang	23	Perhatikan	24
Mengajar	45	Penghianat	38	Bicara	44
Masak	17	Pengangguran	40	Catat	21
Kerja	54	Diberhentikan	45	Pergi	20
Bagus	47	Bandingan	32	Akibat	16
Lindungi	50	Bukan	44	Berita	49
Tepat	23	Hangus	52	Beli	29
Percayakan	33	Kasus	23	Saling	32
Muda	28	Merendahkan	27	Suka	26
Bersahaja	28	Kalah	31	Arah	24
Berkarya	29	Bakar	15	Harus	41
Pikiran	31	Provokasi	43	Tentukan	20
Sukses	56	Gagal	33	Urutan	42
Syukuran	23	Kecewa	36	Sampai	22
Makin	17	Tidak	42	Selamat	51
Harapan	58	Beban	30	Kemampuan	19

Kata Positif	Jumlah	Kata Negatif	Jumlah	Kata Netral	Jumlah
Menguntungkan	46	Dicabut	25	Masa	29
Membuat	30	Dibohongi	43	Undang	19
Menghias	21	Pecah	44	Amanat	45
Kepercayaan	49	Tetapi	46	Penuh	32
Penggiat	44	Berseteru	20	Pembahasan	22
Penghibur	32	Kekerasan	54	Pendahulu	34
Makmur	24	Resah	30	Ungkapan	22
Menjaga	19	Gelisah	46	Presiden	58
Memberi	45	Ragu	33	Mentri	53
Membantu	51	Kerusakan	21	Dpr	54
Menolong	37	Kecelakaan	17	Polisi	49
Meraih	29	Pembunuhan	27	Alam	34
Kebebasan	24	Siksa	45	Kendaraan	50
Kerukunan	35	Hinaan	56	Kapal	23
Kegembiraan	30	Psk	19	Keluarga	20
Indah	26	Sampah	59	Hutan	27
Asri	43	Gersang	45	Janji	56
Elok	23	Curang	47	Jika	48
Meriah	20	Palsu	48	Manusia	44
Pujian	34	Bohong	40	Jual	33
Ajaib	15	Masalah	44	Semua	49
Membela	30	Tertipu	48	Istana	19
Saran	40	Ott	33	Universitas	45
Mencoba	50	Sembarangan	31	Suasana	20
Perlu	30	Dzalim	17	Bangunan	27
Bangga	37	Setan	14	Negara	44
Memimpin	50	Serangan	39	Dilantik	22
Kuliah	60	Diserang	33	Mengurai	36
Tuhan	52	Dibocorkan	17	Analisis	28
Berjuang	36	Merokok	18	Pendapat	48
Utuh	20	Mengganggu	23	Aksi	41
Bersabar	37	Dijepit	13	Dialog	33
Betul	50	Risih	19	Humas	21
Bersih	49	Sinis	29	Media	43
Hayatilah	10	Kontroversi	21	Rahasia	22
Belajar	60	Dendam	14	Hukum	37
Mahasiswa	56	Takpunyaotak	17	Bpk	23
Asyik	24	Dibelah	17	Khusus	31
Kutubuku	29	Fitnah	33	Suara	39
Hukuman	35	Tudugaan	31	Fakultas	46

Kata Positif	Jumlah	Kata Negatif	Jumlah	Kata Netral	Jumlah
Buku	28	Protes	37	Medsos	46
Penerus	30	Ditolak	29	Netizen	49
Generasibaru	32	Membully	24	Wni	26
Dukungpenuh	25	Mati	15	Masa	38
Mulia	19	Dihujat	26	Ngopi	25
Kritik	34	Direkayasa	27	Cerita	27
Tuntutan	33	Kehilanganan	33	Mencari	13
Pendukung	37	Ditunda	21	Bertolak	12
Ikhlas	24	Takjelas	34	Subuh	9
Diraih	33	Lelet	22	Cuitan	45
Pembangunan	44	Menunggu	40	Sidang	45
Hormati	23	Takpeduli	27	Politik	48
Pemuja	20	Dibokong	19	Penasehat	18
Tegar	19	Konflik	40	Revisi	37
Keadilan	39	Merisaukan	26	Mengambil	34
Dispensasi	18	Rasatakut	25	Usaha	26
Integrasi	33	Banyakpolusi	34	Menyikapinya	44
Maju	35	Radikalis	40	Prosedur	32
Pencerahaan	29	Ancaman	41	Mengangkat	16
Mengharumkan	29	Fiktif	33	Menurunkan	27

