

***Opinion Mining Berita Seputar Universitas Sebelas Maret  
Menggunakan Algoritma Naive Bayes***

**SKRIPSI**

**Diajukan untuk memenuhi sebagai persyaratan mendapatkan gelar Strata Satu  
Program Studi Informatika**



**Disusun Oleh:**

**Muhammad Mukhlis Khoirudin**

**M0512042**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS SEBELAS MARET  
SURAKARTA  
2018**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

***Opinion Mining Berita Seputar Universitas Sebelas Maret  
Menggunakan Algoritma Naive Bayes***

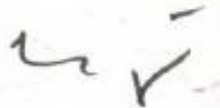
Disusun oleh:

MUHAMMAD MUKHLIS KHOIRUDIN

NIM. M0512042

telah disetujui oleh pembimbing pada tanggal..... 11-01-2018

Pembimbing I



Dr. Wiranto, M.Cs.  
NIP. 196612301993021001

Pembimbing II



Winarno S.Si., M.Eng.  
NIP. 198205202006041001

**HALAMAN PENGESAHAN  
SKRIPSI**

***Opinion Mining Berita Seputar Universitas Sebelas Maret  
Menggunakan Algoritma Naive Bayes***

Disusun oleh:

MUHAMMAD MUKHLIS KHOIRUDIN

NIM. M0512042

telah dipertahankan di hadapan dewan penguji  
pada tanggal 11-01-2018

**Susunan Dewan Penguji**

1. **Dr. Wiranto, M.Kom., M.Cs.** (Ketua)  
NIP. 196612301993021001
2. **Winarno S.Si., M.Eng.** (Sekretaris)  
NIP. 198205202006041001
3. **Ristu Saptono, S.Si., M.T.** (Anggota)  
NIP. 197902102002121001
4. **Dr. Dewi Wisnu W., S.Kom., M.S.** (Anggota)  
NIP. 197810262005012002



**Disahkan Oleh**

Ketua Program Studi Informatika

**Drs. Bambang Harijito, M.App.Sc., PhD**

NIP. 19621130 199103 1 002



# ***Opinion Mining* Berita Seputar Universitas Sebelas Maret Menggunakan Algoritma *Naive Bayes***

**MUHAMMAD MUKHLIS KHOIRUDIN**

Program Studi Informatika Fakultas FMIPA Universitas Sebelas Maret

## **ABSTRAK**

*Opinion mining* atau sentimen analisis adalah cabang dari *text mining* dan merupakan topik yang marak dilakukan saat ini. *Opinion mining* merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Salah satu metode yang bisa digunakan untuk menganalisa dokumen teks adalah klasifikasi. Penelitian ini bertujuan untuk mengklasifikasikan berita Bahasa Indonesia menggunakan *Multinomial Naïve Bayes*.

Untuk mendapatkan hasil yang optimal, penulis mencoba menambahkan fitur seleksi menggunakan *Document Frequency Thresholding (DF-Thresholding)* dan *term weighting* menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*.

Dari hasil penelitian penulis klasifikasi menggunakan *Multinomial Naïve Bayes* mendapatkan akurasi tertinggi dengan rata-rata 90,57%, *Multinomial Naïve Bayes* dengan *DF-Thresholding* mempunyai akurasi 82,92%, serta menggunakan *Multinomial Naïve Bayes* dengan *TF-IDF* memperoleh akurasi 78,5%. Tujuan utama penggunaan seleksi fitur dalam penelitian ini untuk meningkatkan nilai akurasi, namun hasil dalam penelitian ini kurang berpengaruh dari segi akurasi. Penggunaan seleksi fitur bisa mengurangi penggunaan jumlah kompleksitas *term*.

**Kata kunci:** *text mining*, klasifikasi, *multinomial naive bayes*

***Opinion Mining Universitas Sebelas Maret News  
Using Naive Bayes Alghorithm***

**MUHAMMAD MUKHLIS KHOIRUDIN**

*Department of Informatic, Mathematic and Science Faculty Sebelas Maret University*

**ABSTRACT**

*Opinion mining or sentiment analysis is a part of text mining and widespread topic nowadays. Opinion mining is the process of understanding, extracting, and processing textual data automatically to get sentiment information cointained in a sentence. One of the opinion mining method that can be used to analyzed text documents is classification. This research aims to classify Indonesian news into three classes of positive, negative, and neutral using Multinomial Naïve Bayes.*

*To get optimal result, the author tries to add some feature selections using Document Frequency Thresholding (DF-Thresholding) and Term Weighting using Term Frequency-Inverse Document Frequency (TF-IDF).*

*The result showed that the classification using Multinomial Naïve Bayes obtained the highest accuracy with an average 90,57%, Multinomial Naïve Bayes with DF-Thresholding had an accuracy of 82,92% and using Multinomial Naïve Bayes with TF-IDF get an accuracy 78,5%. The actual purpose of using the feature selection in this research to add accuracy value, but the result show less influence in terms of accuracy. Using the selection feature can reduce the use of term complexity.*

**Keyword:** *text mining, classification, multinomial naive bayes*

## MOTTO

*“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”*

*Q.S. Al Baqarah: 286*

*Hai orang-orang yang beriman, jadikanlah sabar dan shalatmu sebagai penolongmu,  
Sesungguhnya Allah beserta orang-orang yang sabar*

*(Al-Baqarah: 153)*

*“Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-  
sungguh (urusan) yang lain,”*

*(Al Insyirah: 7)*

*Barangsiapa belum merasakan pahitnya belajar walau sebentar,  
Ia akan merasakan hinanya kebodohan sepanjang hidupnya.*

*(Imam Syafi'i)*

*Man Jadda Wajada*

*Man Shabara Zhafira*

*Man Sara Darbi Ala Washala*

*(Anonim)*

## PERSEMBAHAN

*Tugas akhir ini saya persembahkan teruntuk  
Kedua orang tuaku, Ibu Sri Hastuti, S.E. dan Bapak Teguh Abdul Hamid, S.Pd.  
Kedua adikku tercinta, Hanifah Rodhiatul Latifah dan Arifin Rois Sholikhin*



## KATA PENGANTAR

Segala puji dan syukur bagi Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “*Opinion Mining* Berita Seputar Universitas Sebelas Maret Menggunakan Algoritma *Naïve Bayes*”.

Dalam menyelesaikan penelitian Tugas Akhir ini, penulis mendapatkan banyak dukungan dan bantuan dari berbagai pihak. Maka dari itu, penulis mengucapkan terimakasih yang tak terhingga kepada :

1. Ibu, bapak, dan adik beserta seluruh anggota keluarga besar yang selalu mendoakan dan selalu memberikan motivasi.
2. Bapak Dr. Wiranto, M.Cs. dan Bapak Winarno S.Si., M.Eng selaku dosen pembimbing I dan pembimbing II atas waktu, ilmu, kesabaran, dan kebaikan yang telah diberikan kepada penulis selama menyelesaikan Tugas Akhir ini.
3. Dosen di Program Studi S1 Informatika, Universitas Sebelas Maret khususnya kepada Bapak Ristu Saptono, S.Si., M.T. yang dengan tulus ikhlas memberikan ilmu dan membimbing selama ini kepada penulis.
4. Data dan Informasi Universitas Sebelas Maret (DI UNS) khususnya Bapak Harmadi Subiatmoko, S.Sos. beserta staff dan teman maganger yang telah membantu penulis untuk mendapatkan data yang digunakan untuk penelitian.
5. Semua teman-teman Informatika angkatan 2012 yang telah membantu penulis dalam pengerjaan tugas akhir ini.
6. Teman seperjuangan Syiar Kegiatan Islam Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sebelas Maret (SKI FMIPA UNS) dan Kesatuan Aksi Mahasiswa Muslim Indonesia (KAMMI) yang telah kebersamai.
7. Sahabat dan komunitas baik di dunia nyata ataupun maya yang tidak bisa saya sebutkan satu persatu yang selalu memberikan inspirasi, motivasi dan hiburan selama proses pengerjaan mengerjakan skripsi.

Surakarta, Januari 2018

Penulis



## DAFTAR ISI

<b>SKRIPSI .....</b>	<b>i</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>iii</b>
<b>ABSTRAK .....</b>	<b>iv</b>
<b>ABSTRACT .....</b>	<b>v</b>
<b>MOTTO .....</b>	<b>vi</b>
<b>PERSEMBAHAN .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI .....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xiv</b>
<b>BAB I.....</b>	<b>1</b>
<b>PENDAHULUAN .....</b>	<b>1</b>
<b>1.1 Latar Belakang.....</b>	<b>1</b>
<b>1.2 Rumusan Masalah.....</b>	<b>3</b>
<b>1.3 Batasan Masalah .....</b>	<b>3</b>
<b>1.4 Tujuan Penelitian .....</b>	<b>4</b>
<b>1.5 Manfaat Penelitian .....</b>	<b>4</b>
<b>1.6 Sistematika Penulisan .....</b>	<b>4</b>
<b>BAB II.....</b>	<b>6</b>
<b>TINJAUAN PUSTAKA.....</b>	<b>6</b>
<b>2.1 Dasar Teori .....</b>	<b>6</b>
<b>2.1.1. Text Mining.....</b>	<b>6</b>

2.1.2.	<i>Opinion Mining</i> .....	6
2.1.3.	<i>Text Preprocessing</i> .....	7
2.1.4.	<i>Feature Selection</i> .....	8
2.1.5.	<i>Document Frequency Thresholding (DF-Thresholding)</i> .....	8
2.1.6.	<i>Term Frequency – Inverse Document Frequency (TF-IDF)</i> .....	9
2.1.7.	<i>Naïve Bayes Classifier</i> .....	9
2.1.8.	<i>Multinomial</i> .....	11
2.2	<b>Penelitian Terkait</b> .....	12
2.3	<b>Rencana Penelitian</b> .....	14
<b>BAB III</b> .....		16
<b>METODOLOGI PENELITIAN</b> .....		16
3.1	<b>Pengumpulan Data</b> .....	16
3.2	<i>Text Preprocessing</i> .....	17
3.3	<i>Document Frequency Thresholding (DF-Thresholding)</i> .....	18
3.4	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	19
3.5	<b>Klasifikasi</b> .....	19
3.6	<b>Evaluasi</b> .....	19
3.7	<i>Kesimpulan dan Saran</i> .....	<b>Error! Bookmark not defined.</b>
<b>BAB IV</b> .....		21
<b>PEMBAHASAN</b> .....		21
4.1	<b>Pengumpulan Data</b> .....	21
4.2	<i>Text Preprocessing</i> .....	24
4.3	<i>Document Frequency Thresholding (DF-Thresholding)</i> .....	27
4.4	<i>Term Frequency – Inverse Document Frequency (TF-IDF)</i> .....	30
4.5	<b>Klasifikasi</b> .....	30
4.5.1	<i>Contoh Perhitungan Multinomial Naive Bayes</i> .....	31

4.5.2 Contoh Perhitungan <i>Multinomial Naïve Bayes TF-IDF</i> .....	33
4.6 Evaluasi.....	35
<b>BAB V</b> .....	50
<b>PENUTUP</b> .....	50
5.1 Kesimpulan .....	50
5.2 Saran .....	51
<b>DAFTAR PUSTAKA</b> .....	52
<b>LAMPIRAN</b> .....	53



**DAFTAR TABEL**

Tabel 2.1 Keterkaitan Penelitian.....	14
Tabel 3.1 Confusion Matrix.....	20
Tabel 4.1 Data Training Penelitian.....	21
Tabel 4.2 Data Testing Penelitian.....	22
Tabel 4.3 Contoh Tabel Berita.....	27
Tabel 4.4 Contoh Perhitungan Term Frequency.....	28
Tabel 4.5 Contoh Perhitungan Document Frequency.....	28
Tabel 4.6 Pengujian DF-Threshold.....	29
Tabel 4.7 Contoh Perhitungan TF-IDF.....	30
Tabel 4.8 Contoh Data Perhitungan Multinomial Naïve Bayes .....	31
Tabel 4.9 Contoh Perhitungan Multinomial Naïve Bayes dengan TF-IDF .....	33
Tabel 4.10 Pengujian DF Memorizing .....	36
Tabel 4.15 Perbandingan Hasil 100 Data Testing .....	40
Tabel 4.16 Perbandingan Hasil 150 Data <i>Testing</i> .....	42
Tabel 4.17 Perbandingan Hasil 200 Data <i>Testing</i> .....	44
Tabel 4.18 Perbandingan Hasil 250 Data <i>Testing</i> .....	46
Tabel 4.11 Perhitungan Klasifikasi 100 Data Testing .....	47
Tabel 4.12 Perhitungan Klasifikasi 150 Data Testing .....	47
Tabel 4.13 Perhitungan Klasifikasi 200 Data Testing .....	48
Tabel 4.14 Perhitungan Klasifikasi 250 Data Testing .....	48

## DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian.....	16
Gambar 4.1 Contoh dokumen kelas positif.....	23
Gambar 4.2 Contoh dokumen kelas negatif.....	23
Gambar 4.3 Contoh dokumen kelas netral.....	24
Gambar 4.4 Proses Case Folding.....	24
Gambar 4.5 Proses Tokenization.....	25
Gambar 4.6 Proses Filtering.....	26
Gambar 4.7 Proses Stemming.....	26
Gambar 4.8 Tampilan Program Testing Data.....	37
Gambar 4.9 Diagram Batang 100 Data Testing MNB.....	38
Gambar 4.10 Diagram Batang 100 Data Testing DF-T + MNB.....	39
Gambar 4.11 Diagram Batang 100 Data Testing TF-IDF + MNB.....	39
Gambar 4.12 Diagram Batang 150 Data Testing MNB.....	41
Gambar 4.13 Diagram Batang 150 Data Testing DF-T + MNB.....	41
Gambar 4.14 Diagram Batang 150 Data Testing TF-IDF + MNB.....	42
Gambar 4.15 Diagram Batang 200 Data Testing MNB.....	43
Gambar 4.16 Diagram Batang 200 Data Testing DF-T + MNB.....	43
Gambar 4.17 Diagram Batang 200 Data Testing TF-IDF + MNB.....	44
Gambar 4.18 Diagram Batang 250 Data Testing MNB.....	45
Gambar 4.19 Diagram Batang 250 Data Testing DF-T + MNB.....	45
Gambar 4.20 Diagram Batang 250 Data Testing TF-IDF + MNB.....	46

## DAFTAR LAMPIRAN

Lampiran 1 Contoh Data Training .....	53
Lampiran 2 Perbandingan Data Training dan Data Testing .....	54
Lampiran 3 Hasil Data Testing 100 dokumen .....	55
Lampiran 4 Hasil Data Testing 150 Dokumen .....	56
Lampiran 5 Hasil Data Testing 200 Dokumen .....	57
Lampiran 6 Hasil Data Testing 250 Dokumen .....	58

