

**ANALISIS DATA ULASAN PEMBELI TERHADAP *SELLER*
PADA SITUS *E-COMMERCE XYZ* MENGGUNAKAN
METODE TF-IDF DAN *NAÏVE BAYES CLASSIFIER***

TUGAS AKHIR



RIZKY AEAMELIA

1142001001

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2018**

**ANALISIS DATA ULASAN PEMBELI TERHADAP *SELLER*
PADA SITUS *E-COMMERCE XYZ* MENGGUNAKAN
METODE TF-IDF DAN *NAÏVE BAYES CLASSIFIER***

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer**



RIZKY AEAMELIA

1142001001

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2018**

HALAMAN PERNYATAAN ORISINALITAS

**Tugas Akhir ini adalah hasil karya saya sendiri, dan
semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**

Nama : Rizky Acamelia

NIM : 1142001001

Tanda Tangan : 

Tanggal : 23 Agustus 2018

HALAMAN PENGESAHAN


Tugas Akhir ini diajukan oleh:

Nama : Rizky Aeamelia
NIM : 1142001001
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Judul Skripsi : Analisis Data Ulasan Pembeli Terhadap *Seller*
Pada Situs *E-Commerce XYZ* Menggunakan Metode
TF-IDF dan *Naïve Bayes Classifier*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer Universitas Bakrie

DEWAN PENGUJI

Pembimbing : Yusuf Lestanto, S.T., M.Sc.



23/8/18

Penguji : Prof. Dr. Hoga Saragih, S.T., M.T.



Penguji : Berkah I. Santoso, S.T., M.T.I.



Ditetapkan di : Jakarta

Tanggal : 23 Agustus 2018

UNGKAPAN TERIMA KASIH

Alhamdulillahirabbil'aalamiin, puji dan syukur kehadiran Allah SWT atas segala rahmat dan karunia-Nya Tugas Akhir yang berjudul “Analisis Data Ulasan Pembeli Terhadap *Seller* Pada Situs *E-Commerce XYZ* Menggunakan Metode TF-IDF dan *Naïve Bayes Classifier*” dapat diselesaikan. Shalawat dan salam penulis haturkan kepada Nabi Besar Muhammad SAW, keluarga dan para sahabatnya yang telah membimbing umatnya ke masa yang terang benderang penuh dengan cahaya iman.

Penyusunan Tugas Akhir ini tidak lepas dari berbagai hambatan dan kesulitan dari awal hingga akhir penyusunan. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada Bapak Yusuf Lestanto, S.T., M.Sc. selaku dosen pembimbing Tugas Akhir yang telah meluangkan waktu serta mengerahkan tenaga dan pikiran untuk membantu proses pengerjaan dan penyusunan Tugas Akhir. Begitu banyak pihak yang turut membantu dan memberikan dukungan, masukan, nasihat, serta doa selama penyusunan Tugas Akhir ini. Oleh karena itu, dengan segala hormat dan kerendahan hati, penulis mengungkapkan rasa terima kasih kepada:

1. Kedua orangtua yang terhormat dan tercinta, serta saudara-saudara tercinta penulis, yang tidak pernah lelah untuk memberikan dukungan, doa, nasihat, masukan, motivasi, dan semangat, serta selalu menjadi pengingat selama ini.
2. Bapak Prof. Dr. Hoga Saragih, S.T., M.T. selaku Ketua Program Studi Informatika Universitas Bakrie dan dosen penguji Tugas Akhir, yang senantiasa memberikan motivasi dan masukan terhadap penyusunan Tugas Akhir.
3. Bapak Berkah I. Santoso, S.T., M.T.I. selaku dosen penguji Tugas Akhir yang senantiasa memberikan motivasi, saran serta perbaikan terhadap penulisan dan penyusunan Tugas Akhir.
4. Fira, Uti, dan Kak Caca yang senantiasa memberikan semangat, nasehat, dan menjadi pendengar yang baik selama proses pengerjaan Tugas Akhir.
5. Teman-teman Informatika 2014: Arisyi Falih Naufal, Imam Sopani, Maissy, Muh Arfandy Wijaya Ak., Muhammad Fauzan Adhim, M. Rahman

Muttaqin, Putri Nurrahmah Arta, Rizka, Safira Alisha Aqista, Shania Isyahrani dan Yossy Alrin, yang selalu membantu, memberikan motivasi, semangat, serta kebersamaan dalam suka dan duka selama 4 tahun masa perkuliahan di Universitas Bakrie.

6. Kakak-kakak Informatika Universitas Bakrie Angkatan 2012 dan 2013 yang telah banyak membagikan pengalaman, memberikan motivasi, bantuan, doa serta semangat selama penyusunan Tugas Akhir.
7. Adik-adik Informatika Universitas Bakrie Angkatan 2015, 2016, dan 2017 yang telah memberikan doa, dukungan, dan semangat selama masa perkuliahan dan penyusunan Tugas Akhir.
8. Sahabat-sahabat tercinta, Halimah, Riryn, Dewi, dan Putri yang senantiasa memberikan semangatnya.

Semoga Allah SWT senantiasa membalas kebaikan dan memberikan keberkahan kepada kita semua. Penulis berharap Tugas Akhir ini dapat memberi informasi yang berguna dan bermanfaat bagi berbagai kalangan bidang pendidikan, khususnya bidang Informatika.

Jakarta, 23 Agustus 2018

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama : Rizky Acamelia
NIM : 1142001001
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Jenis Tugas Akhir : Implementasi Algoritma

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

**Analisis Data Ulasan Pembeli Terhadap *Seller* Pada Situs *E-Commerce XYZ*
Menggunakan Metode TF-IDF dan *Naïve Bayes Classifier***

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 2018

Yang menyatakan



Rizky Acamelia

**ANALISIS DATA ULASAN PEMBELI TERHADAP SELLER PADA
SITUS E-COMMERCE XYZ MENGGUNAKAN METODE
TF-IDF DAN NAÏVE BAYES CLASSIFIER**

Rizky Aeamelia

ABSTRAK

Popularitas *e-commerce* terus mengalami peningkatan yang cukup signifikan. Sehingga penelitian mengenai *e-commerce* menjadi topik yang menarik untuk diteliti. Salah satunya yang cukup populer adalah situs *e-commerce* XYZ. Banyak pertimbangan yang dilakukan oleh calon pembeli sebelum melakukan pembelian produk. Hal seperti ini sering ditemukan pada *online marketplace* dimana banyaknya kriteria *seller* yang memberikan berbagai pilihan produk sehingga calon pembeli semakin sulit untuk memutuskan pembeliannya. Masalah ini dapat dibantu dengan fitur ulasan pembeli yang dikenal dengan *Online Customer Review* (OCR) yang membantu calon pembeli untuk menentukan keputusan pembelian berdasarkan ulasan yang telah diberikan oleh konsumen sebelumnya. Berdasarkan hal tersebut, peneliti tertarik untuk melakukan penelitian terkait OCR pada situs *e-commerce* XYZ dalam rangka menentukan *seller* yang memiliki pelayanan terbaik berdasarkan jumlah ulasan positif terbanyak dengan menerapkan metode TF-IDF dan *Naïve Bayes Classifier*. Dalam penelitian ini, peneliti mengumpulkan data ulasan pembeli sebanyak 1500 ulasan yang diambil dari 10 *seller* yang termasuk kategori *Kesehatan dan Kecantikan*. Dari 1500 ulasan, kemudian dibagi dalam dua *subset* data, yaitu *data training* sebanyak 1000, dan *data testing* sebanyak 500. Dari hasil proses klasifikasi dengan menggunakan 500 *data testing* dengan masing-masing 50 ulasan pembeli untuk setiap *seller*, ditunjukkan bahwa *seller* dengan nama FAM memiliki jumlah ulasan positif terbanyak sebesar 49 ulasan positif. Hasil pengukuran performa terhadap proses klasifikasi ulasan pembeli ini juga menghasilkan nilai *accuracy* sebesar 97,60%, *precision* 98,83%, *recall* 97,69%, dan *f-measure* sebesar 98,26%. Hal ini membuktikan bahwa proses pengklasifikasian dengan menerapkan tahap *preprocessing*, TF-IDF, dan penggunaan metode *Naïve Bayes Classifier* cukup baik dalam mengklasifikasikan data ulasan pembeli.

Kata Kunci: *Online Customer Review* (OCR), *Preprocessing*, TF-IDF, *Naïve Bayes Classifier*

ANALYSIS OF DATA REVIEW OF BUYERS ON SELLERS ON E-COMMERCE XYZ SITES USING TF-IDF AND NAÏVE BAYES CLASSIFIER METHODS

Rizky Aeamelia

ABSTRACT

The popularity of e-commerce continues to experience a significant increase. E-Commerce research becomes one of a topic that interesting to investigate. One of the most popular is e-commerce XYZ sites. Many considerations are made by prospective buyers before purchasing a product. Things like this are often found in the online marketplace where many sellers criteria that provide a wide selection of products so that potential buyers are face difficulty to decide on the purchase. This problem can be helped with a feature of buyer's review known as Online Customer Review (OCR), which helps potential buyers to determine purchasing decisions based on the reviews given by previous customers. This Paper will conduct research related to the OCR on e-commerce sites in order to determine the seller who has the best service based on the highest number of positive reviews by applying TF-IDF and Naïve Bayes classifier. In this study, this research collected 1500 buyer's review data taken from 10 sellers including the Health and Beauty category. From 1500 reviews, it'll divided into two subset of data, namely training data as many as 1000, and testing data as many as 500. From the results of the classification process using 500 testing data with each of the 50 buyer reviews for each seller, it was shown that the seller with the name FAM had the most positive reviews for 49 positive reviews. The performance measurement results on the buyer review classification process also produce accuracy values of 97,60%, precision 98,83%, recall 97,69%, and f-measure of 98,26%. This proves that the classification process by applying the preprocessing, TF-IDF, and the use of the Naïve Bayes Classifier method is good enough to classify buyer review data.

Keywords: Online Customer Review (OCR), Preprocessing, TF-IDF, Naïve Bayes Classifier

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN	iii
UNGKAPAN TERIMA KASIH	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	vi
ABSTRAK	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR <i>LISTING CODE</i>	xiv
DAFTAR RUMUS	xv
DAFTAR SINGKATAN	xvi
DAFTAR LAMPIRAN	xvii
 BAB I	 1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
 BAB II	 7
TINJAUAN PUSTAKA	7
2.1 Penelitian Terkait	7
2.2 <i>Text Mining</i>	16
2.3 Analisis Sentimen	17
2.3.1 Tingkatan Klasifikasi Sentimen	17
2.3.2 Teknik Klasifikasi Sentimen	19

2.3.3	Tahapan Analisis Sentimen.....	22
2.4	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	23
2.5	<i>Naïve Bayes Classifier</i>	25
2.5.1	<i>Multinomial Naïve Bayes Classifier</i>	26
2.6	Pengukuran Performa.....	28
BAB III.....		32
METODOLOGI PENELITIAN		32
3.1	Tahapan Penelitian.....	32
3.1.1	Studi Literatur.....	32
3.1.2	Rumusan Masalah.....	32
3.1.3	Melakukan Penelitian	33
3.1.4	Penyusunan Laporan.....	33
3.2	Kerangka Penelitian.....	34
3.2.1	Pengumpulan Data.....	35
3.2.2	<i>Dataset Ulasan Pembeli</i>	36
3.2.3	<i>Preprocessing</i>	37
3.2.4	Seleksi Fitur (TF-IDF)	41
3.2.5	Pembagian Data.....	41
3.2.6	Proses <i>Naïve Bayes Classifier</i>	42
3.2.7	Pengukuran Performa Klasifikasi.....	50
3.3	Instrumen Penelitian	50
3.3.1	Perangkat Keras (<i>Hardware</i>)	50
3.3.2	Perangkat Lunak (<i>Software</i>).....	50
3.4	Rencana Kegiatan Penelitian.....	51
BAB IV		52
IMPLEMENTASI DAN HASIL PENELITIAN.....		52
4.1	Pengumpulan Data.....	52
4.2	<i>Dataset Ulasan Pembeli</i>	52
4.3	Hasil <i>Preprocessing</i>	53
4.4	Seleksi Fitur (TF-IDF)	56
4.5	Klasifikasi <i>Naïve Bayes</i>	57

4.6 Pengukuran Performa.....	62
BAB V.....	65
KESIMPULAN DAN SARAN	65
5.1 Kesimpulan.....	65
5.2 Saran	65
DAFTAR PUSTAKA	67
LAMPIRAN	76

DAFTAR GAMBAR

Gambar 2. 1 Tingkatan Level Analisis Sentimen.....	18
Gambar 2. 2 Teknik Klasifikasi Sentimen	19
Gambar 2. 3 Alur <i>Supervised Learning</i>	20
Gambar 2. 4 Proses Analisis Sentimen pada Produk.....	23
Gambar 2. 5 <i>K-fold Cross Validation Pseudocode</i>	29
Gambar 3. 1 Tahapan Penelitian.....	32
Gambar 3. 2 Kerangka Penelitian.....	34
Gambar 3. 3 Hasil <i>Scraping</i>	35
Gambar 3. 4 Contoh Sentimen Positif	36
Gambar 3. 5 Contoh Sentimen Negatif	36
Gambar 3. 6 <i>Flowchart</i> Proses <i>Stopword Removal</i>	40
Gambar 3. 7 <i>Pseudocode Stopword Removal</i>	40
Gambar 3. 8 <i>Flowchart</i> Data <i>Training</i>	43
Gambar 3. 9 <i>Flowchart</i> Klasifikasi Data <i>Testing</i>	47
Gambar 4. 1 Potongan Hasil TF-IDF.....	57
Gambar 4. 2 Pembagian Data.....	57
Gambar 4. 3 <i>Pseudocode</i> Algoritma <i>Naive Bayes (Multinomial): Training</i>	58
Gambar 4. 4 <i>Pseudocode</i> Algoritma <i>Naive Bayes (Multinomial): Testing</i>	60
Gambar 4. 5 Hasil Klasifikasi pada Data <i>Testing</i> untuk Setiap <i>Seller</i>	62

DAFTAR TABEL

Tabel 2. 1 Rangkuman Penelitian Terkait	12
Tabel 2. 2 <i>Confusion Matrix</i>	30
Tabel 3. 1 Contoh Proses <i>Cleansing</i>	37
Tabel 3. 2 Contoh Proses <i>Case Folding</i>	38
Tabel 3. 3 Contoh Proses <i>Stemming</i>	38
Tabel 3. 4 Contoh Proses <i>Stopword Removal</i>	41
Tabel 3. 5 Contoh Pelabelan Data <i>Review Pembeli</i>	43
Tabel 3. 6 Frekuensi Kemunculan Kata pada Sampel Data Latih.....	44
Tabel 3. 7 Model Probabilistik	46
Tabel 3. 8 Hasil <i>Preprocessing</i> pada Sampel Data Uji.....	48
Tabel 3. 9 Hasil Perhitungan Peluang Setiap Kata pada Sampel Data Uji	48
Tabel 3. 10 Rencana Kegiatan Penelitian	51
Tabel 4. 1 Potongan <i>Dataset</i> Ulasan Pembeli	52
Tabel 4. 2 Contoh Penambahan Kata Pada Daftar <i>Stopwords</i>	55
Tabel 4. 3 Potongan <i>Raw Data</i> Hasil <i>Preprocessing</i>	56
Tabel 4. 4 Perbandingan Data <i>Predicted</i> dan <i>Actual</i>	63
Tabel 4. 5 Nilai <i>Confusion Matrix</i> Analisis Sentimen.....	63

DAFTAR LISTING CODE

<i>Listing 4. 1 Potongan Source Code Cleansing dan Tokenizing</i>	53
<i>Listing 4. 2 Potongan Source Code Case Folding</i>	54
<i>Listing 4. 3 Potongan Source Code Stemming</i>	54
<i>Listing 4. 4 Potongan Source Code Stopword Removal</i>	56
<i>Listing 4. 5 Parameter atau Query</i>	56
<i>Listing 4. 6 Source Code Menghitung Prior</i>	58
<i>Listing 4. 7 Source Code Menghitung Conditional Probability</i>	59
<i>Listing 4. 8 Source Code Mencari Nilai C_{MAP}</i>	60
<i>Listing 4. 9 Source Code Menentukan Kelas Sentimen</i>	61
<i>Listing 4. 10 Source Code Perhitungan Jumlah Sentimen pada Data Testing</i>	61

DAFTAR RUMUS

Persamaan (2. 1) <i>Term Frequency</i> (TF)	23
Persamaan (2. 2) <i>Inverse Document Frequency</i> (IDF)	24
Persamaan (2. 3) <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF).....	24
Persamaan (2. 4) Perhitungan umum NBC	25
Persamaan (2. 5) Penyederhanaan perhitungan umum NBC	26
Persamaan (2. 6) Perhitungan <i>prior</i> pada NBC	26
Persamaan (2. 7) Perhitungan <i>posterior</i> pada NBC.....	26
Persamaan (2. 8) Perhitungan <i>conditional probability</i> pada MNB	27
Persamaan (2. 9) Perhitungan umum MAP pada MNB	27
Persamaan (2. 10) Penyederhanaan perhitungan MAP pada MNB.....	28
Persamaan (2. 11) Penyederhanaan (II) perhitungan MAP pada MNB.....	28
Persamaan (2. 12) <i>Accuracy</i>	30
Persamaan (2. 13) <i>Precision</i>	30
Persamaan (2. 14) <i>Recall</i>	31
Persamaan (2. 15) <i>F-measure</i>	31

DAFTAR SINGKATAN

DFWNB	<i>Deep Feature Weighting Naive Bayes</i>
IDF	<i>Inverse Document Frequency</i>
MAP	<i>Maximum a Posterior</i>
MNB	<i>Multinomial Naïve Bayes</i>
NBC	<i>Naïve Bayes Classifier</i>
NLP	<i>Natural Language Processing</i>
SVM	<i>Support Vector Machine</i>
TF	<i>Term Frequency</i>
WEKA	<i>Waikato Environment for Knowledge</i>

DAFTAR LAMPIRAN

Lampiran 1: Potongan *Dataset* Ulasan Pembeli

Lampiran 2: Daftar stopWord.txt