

ABSTRAK

Twitter merupakan sebuah aplikasi *social networking* yang memungkinkan *user*nya untuk dapat mengirimkan pesan pada waktu yang bersamaan. Data yang diambil melalui Twitter dapat dijadikan sebagai media penelitian dalam analisis sentimen. Analisis sentimen adalah mekanisme pengelompokan kalimat ke dalam beberapa kategori, antara lain kategori positif, negatif, ataupun netral. *Naïve Bayes Classifier* merupakan salah satu teknik klasifikasi yang digunakan dalam mengkategorikan sebuah kalimat menjadi positif, negatif, atau netral. Pada pembuatan aplikasi analisis sentimen ini terdapat dua data utama yang berperan penting untuk memastikan validitas analisis sentimen, yaitu *data training* dan *data testing*. Pembuatan *data training* akan melalui beberapa tahapan diantaranya adalah *pre-processing*, *mutual information*, dan klasifikasi. *Data testing* dengan tingkat akurasi yang tinggi akan menyimpulkan bahwa aplikasi analisis sentimen ini sudah berjalan dengan baik. Sedangkan jika *data testing* menghasilkan tingkat akurasi yang rendah, maka dapat disimpulkan bahwa sistem harus *training* ulang.

Kata Kunci: analisis sentimen, *pre-processing*, *mutual information*, dan *Naïve Bayes*.

ABSTRACT

Twitter is a social networking application that allows its user to be able to send messages at the same time. The data which collected via Twitter can be used as a media research in the sentiment analysis. Sentiment analysis is a mechanism of grouping sentences into several categories, such as positive, negative, or neutral. Naïve Bayes Classifier is one of the technique used in classification to categorize a sentence into positive, negative, or neutral. In the process of making this application, there are two main data which play an important role to ensure the validity of sentiment analysis, which are data training and data testing. Data training is created through several stages, there are pre-processing, mutual information, and classification. Data testing with high accuracy will conclude that this sentiment analysis application has worked well. However, if data testing given the result with low accuracy, it will conclude that the system has to be trained again.

Keywords: sentiment analysis, pre-processing, mutual information, and Naïve Bayes Classifier

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS LAPORAN PENELITIAN	ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN	iii
PRAKATA	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
DAFTAR PROGRAM	xv
DAFTAR NOTASI/LAMBANG	xvi
DAFTAR SINGKATAN	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan	2
1.4. Batasan Masalah	2
1.5. Sistematika Penyajian	3
BAB II KAJIAN TEORI	4
2.1. Analisis Sentimen	4
2.2. Pre-processing	4
2.2.1. Tokenizing	6
2.2.2. Formalization	6
2.2.3. Stopping	6
2.2.4. N-Grams	7
2.3. Mutual Information	9
2.4. Naïve Bayes Classifier	10
2.4.1. Laplacian Smoothing	13
2.5. Twitter API	15
BAB III ANALISIS DAN DISAIN	16
3.1. Analisis	16
3.1.1. Contoh Penerapan Analisis	17
3.1.1.1. Tokenizing	18
3.1.1.2. Formalization	18
3.1.1.3. Stopping	19
3.1.1.4. N-Grams	19
3.1.1.5. Mutual Information	22
3.1.1.6. Proses Pembuatan Basis Term untuk Data Training	23
3.1.1.7. Proses Training	24
3.1.1.8. Proses Testing	27
3.2. Gambaran Keseluruhan	28
3.2.1. Persyaratan Antarmuka Eksternal	28
3.2.2. Antarmuka dengan Pengguna	28

3.2.3.	Antarmuka Perangkat Keras	28
3.2.4.	Antarmuka Perangkat Lunak.....	28
3.3.	Disain Perangkat Lunak	29
3.3.1.	Pemodelan Perangkat Lunak.....	29
3.3.1.1.	Arsitektur Perangkat Lunak.....	29
3.3.1.1.1	Arsitektur Proses Pemilihan Term	29
3.3.1.1.2	Arsitektur Proses Pembuatan Data Training	30
3.3.1.1.3	Arsitektur Sistem Analisis Sentimen.....	31
3.3.1.2.	Use Case	32
3.3.1.3.	Use Case Skenario	32
3.3.1.3.1	Use Case Manage Data Training.....	32
3.3.1.3.2	Use Case Preprocessing.....	33
3.3.1.3.3	Use Case Mutual Information	33
3.3.1.3.4	Use Case Naïve Bayes.....	34
3.3.1.3.5	Use Case Login	34
3.3.1.3.6	Use Case Twitter API.....	35
3.3.1.3.7	Use Case Display Result	35
3.3.1.3.8	Use Case Search Tweets.....	36
3.3.1.3.9	Use Case Update Data Training	36
3.3.1.4.	Activity Diagram	37
3.3.1.4.1	Activity Diagram Manage Data Training.....	37
3.3.1.4.2	Activity Diagram Processing File	38
3.3.1.4.3	Activity Diagram Mutual Information	40
3.3.1.4.4	Activity Diagram Pembuatan Data Training.....	41
3.3.1.4.5	Activity Diagram Login	42
3.3.1.4.6	Activity Diagram Testing	43
3.3.1.4.7	Activity Diagram Searching	44
3.3.1.4.8	Activity Diagram Update Data Training	45
3.3.2.	Disain Antarmuka	46
3.3.2.1.	Rancangan Halaman Default	46
3.3.2.2.	Rancangan Halaman Login.....	46
3.3.2.3.	Rancangan Halaman Data Training.....	47
3.3.2.4.	Rancangan Halaman Pre-processing	48
3.3.2.5.	Rancangan Halaman Mutual Information.....	48
3.3.2.6.	Rancangan Halaman Basis Term.....	49
3.3.2.7.	Rancangan Halaman Authorize	49
3.3.2.8.	Rancangan Halaman Sentiment Analysis Charting.....	50
3.3.2.9.	Rancangan Halaman Sentiment Analysis Details.....	51
3.3.2.10.	Rancangan Halaman Searching.....	51
3.3.2.11.	Rancangan Halaman Searching Result.....	52
3.3.2.12.	Rancangan Halaman Update Data Training	52
BAB IV	PENGEMBANGAN PERANGKAT LUNAK	53
4.1.	Implementasi Class / Modul	53
4.1.1.	Class Preprocessing.....	53
4.1.2.	Class Selection	58
4.1.3.	Class Classification	59
4.1.4.	Static Class	60

4.1.5.	Main Class.....	61
4.2.	Implementasi Antarmuka	62
4.2.1.	Halaman Default	62
4.2.2.	Halaman Login.....	63
4.2.3.	Halaman Data Training.....	64
4.2.4.	Halaman Tweets.....	64
4.2.5.	Halaman Tokenizing	65
4.2.6.	Halaman Formalization.....	66
4.2.7.	Halaman Stopping.....	66
4.2.8.	Halaman N-Grams	67
4.2.9.	Halaman Mutual Information.....	67
4.2.10.	Halaman Basis Term	68
4.2.11.	Halaman Authorize.....	68
4.2.12.	Halaman Sentiment Analysis Charting	69
4.2.13.	Halaman Sentiment Analysis Details	70
4.2.14.	Halaman Sentiment Analysis Searching	70
4.2.15.	Halaman Sentiment Analysis Searching Result	71
4.2.16.	Halaman Update Data Training.....	71
BAB V TESTING DAN EVALUASI SISTEM		73
5.1.	Halaman Default	73
5.2.	Halaman Login.....	73
5.3.	Halaman Data Training.....	74
5.4.	Halaman Tweets.....	74
5.5.	Halaman Tokenizing	74
5.6.	Halaman Formalization.....	74
5.7.	Halaman Stopping.....	75
5.8.	Halaman N-Grams	75
5.9.	Halaman Mutual Information.....	75
5.10.	Halaman Basis Term	75
5.11.	Halaman Authorize	76
5.12.	Halaman Sentiment Analysis Charting	76
5.13.	Halaman Sentiment Analysis Details.....	77
5.14.	Halaman Searching	77
5.15.	Halaman Update Data Training	77
5.16.	Hasil Testing Data Training dengan Memperhitungkan Formal Words....	78
5.17.	Hasil Testing Data Training dengan Memperhitungkan Stopwords.....	78
5.18.	Hasil Testing Tingkat Akurasi Perhitungan Naïve Bayes.....	78
5.19.	Hasil Testing Tingkat Persentase dengan Memperhitungkan Keywords ..	80
5.20.	Hasil Testing Tingkat Persentase Optimal pada Class Positif	81
5.21.	Hasil Testing Tingkat Persentase Optimal pada Class Negatif.....	82
5.22.	Hasil Testing Tingkat Persentase Optimal pada Class Netral.....	82
5.23.	Hasil Testing Tingkat Persentase Optimal pada Sistem Analisis Sentimen	83
BAB VI KESIMPULAN DAN SARAN		84
6.1.	Kesimpulan	84
6.2.	Saran.....	85
DAFTAR PUSTAKA		86

DAFTAR GAMBAR

Gambar 3.1 Contoh Data pada File Teks	17
Gambar 3.2 Kalimat-kalimat yang Telah Dipilah ke Dalam File Teks	17
Gambar 3.3 Kalimat-kalimat yang Telah Dilakukan Proses Tokenizing	18
Gambar 3.4 Kalimat-kalimat yang Telah Dilakukan Proses Formalization	18
Gambar 3.5 Kalimat-kalimat yang Telah Dilakukan Proses Stopping	19
Gambar 3.6 Arsitektur Proses Pemilihan Term	29
Gambar 3.7 Arsitektur Proses Pembuatan Data Training	30
Gambar 3.8 Arsitektur Sistem Analisis Sentimen.....	31
Gambar 3.9 Use Case Diagram	32
Gambar 3.10 Activity Diagram Manage Data Training.....	37
Gambar 3.11 Activity Diagram Processing File	38
Gambar 3.12 Activity Diagram Mutual Information	40
Gambar 3.13 Activity Diagram Pembuatan Data Training.....	41
Gambar 3.14 Activity Diagram Login	42
Gambar 3.15 Activity Diagram Testing.....	43
Gambar 3.16 Activity Diagram Searching.....	44
Gambar 3.17 Activity Diagram Update Data Training.....	45
Gambar 3.18 Rancangan Halaman Default.....	46
Gambar 3.19 Rancangan Halaman Login	46
Gambar 3.20 Rancangan Halaman Data Training	47
Gambar 3.21 Rancangan Halaman Pre-processing.....	48
Gambar 3.22 Rancangan Halaman Mutual Information	48
Gambar 3.23 Rancangan Halaman Basis Term	49
Gambar 3.24 Rancangan Halaman Authorize.....	49
Gambar 3.25 Rancangan Halaman Sentiment Analysis Charting	50
Gambar 3.26 Rancangan Halaman Sentiment Analysis Details	51
Gambar 3.27 Rancangan Halaman Searching.....	51
Gambar 3.28 Rancangan Halaman Searching Result	52
Gambar 3.29 Rancangan Halaman Update Data Training.....	52
Gambar 4.1 Class Diagram Sentiment Analysis System MCU	53
Gambar 4.2 Class Preprocessing.....	53
Gambar 4.3 Class Selection	58
Gambar 4.4 Class Classification	59
Gambar 4.5 Static Class	60
Gambar 4.6 Main Class.....	61
Gambar 4.7 Halaman Default	62
Gambar 4.8 Halaman Login.....	63
Gambar 4.9 Halaman Login Error	63
Gambar 4.10 Halaman Data Training	64
Gambar 4.11 Halaman Tweets.....	64
Gambar 4.12 Halaman Tokenizing	65
Gambar 4.13 Halaman Formalization	66
Gambar 4.14 Halaman Stopping	66

Gambar 4.15 Halaman N-Grams.....	67
Gambar 4.16 Halaman Mutual Information.....	67
Gambar 4.17 Halaman Basis Term	68
Gambar 4.18 Halaman Authorize	68
Gambar 4.19 Halaman Login Twitter	69
Gambar 4.20 Halaman Sentiment Analysis Charting	69
Gambar 4.21 Halaman Sentiment Analysis Details.....	70
Gambar 4.22 Halaman Sentiment Analysis Searching	70
Gambar 4.23 Halaman Sentiment Analysis Searching Result	71
Gambar 4.24 Halaman Update Data Training.....	71
Gambar 4.25 Notifikasi Data Training Telah Diperbaharui	72
Gambar 5.1 Grafik Hasil Testing Tingkat Persentase dengan Memperhitungkan Keywords	81

DAFTAR TABEL

Tabel 2.1 Emoticons.....	5
Tabel 2.2 Hasil Bi-gram.....	7
Tabel 2.3 Model n-gram dari Hasil Bigram	8
Tabel 2.4 Sample Data Mutual Information	9
Tabel 2.5 Contoh Sample Data NBC	11
Tabel 2.6 Nilai Probabilistik pada Atribut Age	13
Tabel 2.7 Laplacian Smoothing pada Atribut Age.....	14
Tabel 2.8 Nilai Probabilistik pada Atribut Age Setelah Dilakukan Smoothing	14
Tabel 3.1 Kalimat-kalimat yang Telah Dilakukan Proses N-Grams.....	19
Tabel 3.2 Model N-gram pada File NgramsPositif.txt.....	20
Tabel 3.3 Model N-gram pada File NgramsNegatif.txt	21
Tabel 3.4 Model N-gram pada File NgramsNetral.txt	22
Tabel 3.5 Hasil Perhitungan Mutual Information pada Class Positif dari File MIPositif.txt	22
Tabel 3.6 Hasil Perhitungan Mutual Information pada Class Negatif dari File MINegatif.txt.....	23
Tabel 3.7 Hasil Perhitungan Mutual Information pada Class Netral dari File MINetral.txt.....	23
Tabel 3.8 Hasil Perhitungan Mutual Information yang Telah Diurutkan pada Class Positif	23
Tabel 3.9 Hasil Perhitungan Mutual Information yang Telah Diurutkan pada Class Negatif.....	23
Tabel 3.10 Hasil Perhitungan Mutual Information yang Telah Diurutkan pada Class Netral.....	24
Tabel 3.11 Basis Term untuk Data Training.....	24
Tabel 3.12 Hasil Kalimat yang Diubah Menjadi Format Naïve Bayes.....	25
Tabel 3.13 Hasil Nilai Probabilistik untuk Setiap Atribut	25
Tabel 3.14 Laplacian Smoothing pada Atribut Kasih.....	26
Tabel 3.15 Laplacian Smoothing pada Atribut Tuhan	26
Tabel 3.16 Laplacian Smoothing pada Atribut Cape	26
Tabel 3.17 Laplacian Smoothing pada Atribut Lelah	26
Tabel 3.18 Laplacian Smoothing pada Atribut Selamat Pagi	26
Tabel 3.19 Laplacian Smoothing pada Atribut Pagi	26
Tabel 3.20 Hasil Data Training.....	26
Tabel 3.21 Basis Term untuk Data Testing.....	27
Tabel 5.1 Blackbox Testing Halaman Default	73
Tabel 5.2 Blackbox Testing Halaman Login	73
Tabel 5.3 Blackbox Testing Halaman Data Training.....	74
Tabel 5.4 Blackbox Testing Halaman Tweets	74
Tabel 5.5 Blackbox Testing Halaman Tokenizing.....	74
Tabel 5.6 Blackbox Testing Halaman Formalization	74
Tabel 5.7 Blackbox Testing Halaman Stopping	75
Tabel 5.8 Blackbox Testing Halaman N-Grams	75

Tabel 5.9 Blackbox Testing Halaman Mutual Information	75
Tabel 5.10 Blackbox Testing Halaman Basis Term.....	75
Tabel 5.11 Blackbox Testing Halaman Authorize	76
Tabel 5.12 Blackbox Testing Halaman Sentiment Analysis Charting.....	76
Tabel 5.13 Blackbox Testing Halaman Sentiment Analysis Details	77
Tabel 5.14 Blackbox Testing Halaman Searching	77
Tabel 5.15 Blackbox Testing Halaman Update Data Training	77
Tabel 5.16 Hasil Testing Data Training dengan Memperhitungkan Formal Words .	78
Tabel 5.17 Hasil Testing Data Training dengan Memperhitungkan Stopwords.....	78
Tabel 5.18 Hasil Testing Tingkat Akurasi Perhitungan Naïve Bayes	79
Tabel 5.19 Hasil Testing Tingkat Persentase dengan Memperhitungkan Keywords	80
Tabel 5.20 Hasil Testing Tingkat Persentase Optimal pada Class Positif	81
Tabel 5.21 Hasil Testing Tingkat Persentase Optimal pada Class Negatif.....	82
Tabel 5.22 Hasil Testing Tingkat Persentase Optimal pada Class Netral.....	82
Tabel 5.23 Hasil Testing Tingkat Persentase Optimal pada Sistem Analisis Sentimen.....	83

DAFTAR PROGRAM

Kode Program 4.1 Pseudocode Tokenizing	55
Kode Program 4.2 Pseudocode Formalization.....	55
Kode Program 4.3 Pseudocode Stopping.....	56
Kode Program 4.4 Pseudocode N-Grams	57
Kode Program 4.5 Pseudocode Mutual Information.....	59

DAFTAR NOTASI/LAMBANG

Jenis	Notasi/Lambang	Nama	Arti
Use Case Diagram		Actor	Peran pengguna sistem.
Use Case Diagram		Use Case	Operasi – operasi yang dilakukan oleh aktor.
Use Case Diagram		Asosiation	Penghubung antara aktor dengan use case.
Use Case Diagram		System Boundary	Struktural sistem.
Activity Diagram		Initial State	Proses Mulai
Activity Diagram		Final State	Proses Berhenti.
Activity Diagram		Action Node	Proses yang akan dilakukan.
Activity Diagram		Decision Node	Kondisi percabangan.
Activity Diagram		Control Flow	Alur Proses.

DAFTAR SINGKATAN

1. API : *Application Programming Interface*
2. HTTP : *Hypertext Transfer Protocol*
3. XML : *Extensible Markup Language*
4. JSON : *Javascript Object Notation*
5. URL : *Uniform Resource Locator*