ABSTRAK

Teknologi web dan jumlah pengguna internet berkembang dengan sangat pesat. Perkembangan media *online* mendorong munculnya informasi tekstual yang tidak terbatas. Twitter merupakan salah satu media *online* yang memungkinkan pengguna untuk mengirimkan pesan. Pesan yang ada pada Twitter dapat digunakan untuk menganalisis sentimen konsumen terhadap suatu produk. Salah satu cara untuk melakukan analisis sentimen yaitu *Naïve Bayes Classifier (NBC). NBC* adalah salah satu teknik klasifikasi yang digunakan untuk mengklasifikasi kalimat menjadi positif, negatif, ataupun netral. Untuk membuat sebuah aplikasi analisis sentimen diperlukan *data training* dan *data testing*. Data yang digunakan yaitu *tweets* yang berhubungan dengan beberapa *provider* telekomunikasi di Indonesia. Tahap-tahap pembuatan *data training* dimulai dari pengambilan data pada Twitter, *pre processing*, *manual judgment*, dan yang terakhir *selection*. *Data training* yang digunakan ada 1457 *tweet* yang diambil antara 15 Mei 2015 sampai tanggal 22 Mei 2015. Ada beberapa *data testing* yang digunakan salah satunya yaitu dengan menggunakan *data training* itu sendiri. *Analisis* sentimen pada aplikasi memiliki tingkat keakuratan sekitar 70%. Persentase *tweets* pada beberapa *provider* telekomunikasi cenderung lebih banyak *tweets* negatif dan netral daripada positif.

Kata Kunci: Twitter, analisis sentimen, *pre-processing*, *Naïve Bayes*.

ABSTRACT

Web technology and the number of Internet users growing very rapidly. The development of online media to encourage the emergence of textual information which is not limited. Twitter is one of the online media that allows users to send messages. Existing messages on Twitter can be used to analyze consumer sentiment to a product. One way to perform sentiment analysis that Naïve Bayes Classifier (NBC). NBC is one of the classification techniques are used to classify the sentence be positive, negative, or neutral. To create an application sentiment analysis required training data and data testing. The data used are tweets related to multiple telecommunications providers in Indonesia. The stages of manufacture of the training data starting from the collection of data on Twitter, pre-processing, manual judgment, and the final selection. Training data are used there in 1457 tweets were taken between May 15, 2015 until May 22, 2015. There are some data testing that used one of them is by using the training data itself. Sentiment analysis on the application have the accuracy level of about 70%. The percentage of tweets in some telecom providers tend to be more negative tweets and neutral rather than positive.

Keywords : Twitter , sentiment analysis , pre - processing , Naïve Bayes .

DAFTAR ISI

PERNYATAAN ORISINALITAS LAPORAN PENELITIAN ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN iii
PRAKATA iv
ABSTRAK v
ABSTRACT vi
DAFTAR ISI vii
DAFTAR GAMBAR xii
DAFTAR TABEL xiv
DAFTAR NOTASI/LAMBANG xvi
BAB 1 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 2 KAJIAN TEORI 4
2.1. Data Mining 4
2.2. Analisis Sentimen 4
2.3. Konsep Klasifikasi 5
2.4. Mutual Information 5
2.5. Naïve Bayes Classifier 6
2.6. Tokenizing 7
2.7. Formalization 8
2.8. Stopping 8
2.9. N-Grams 8
2.10. Twitter & Twitter API 8
2.11. Evaluasi Tingkat Keberhasilan 10
BAB 3 ANALISIS DAN DISAIN 12
3.1. Analisis 12

3.1.1. Contoh Penerapan Analisis 12
3.1.1.1. Tokenizing 13
3.1.1.2. Formalization 14
3.1.1.3. Stopping 14
3.1.1.4. N-Grams 15
3.1.1.5. Mutual Information 18
3.1.1.6. Proses Training Naïve Bayes 19
3.1.1.7. Testing 21
3.2. Gambaran Keseluruhan 22
3.2.1. Persyaratan Antarmuka Eksternal 22
3.2.2. Antarmuka dengan Pengguna 22
3.2.3. Antarmuka Perangkat Keras 22
3.2.3.1. Spesifikasi Antarmuka Perangkat Keras Saat Pengembangan 23
3.2.3.2. Spesifikasi Antarmuka Perangkat Keras *End-User* 23
3.2.4. Antarmuka Perangkat Lunak 23
3.3. Disain Perangkat Lunak 23
3.3.1. Pemodelan Perangkat Lunak 23
3.3.1.1. Arsitektur Perangkat Lunak 24
3.3.1.1.1 Arsitektur Pembuatan Data Training 24
3.3.1.1.2 Arsitektur Pemilihan Term 25
3.3.1.1.3 Arsitektur Analisis Sentimen 27
3.3.1.2. Use Case 28
3.3.1.3. Use Case Skenario 28
3.3.1.3.1 Use Case Do Manual Judgment 28
3.3.1.3.2 Use Case Preprocessing 29
3.3.1.3.3 Use Case Calculate Mutual Information 29
3.3.1.3.4 Use Case Get Tweets 29
3.3.1.3.5 Use Case Do Classification 30
3.3.1.4. Activity Diagram 31
3.3.1.4.1 Activity Diagram Get Tweets 31
3.3.1.4.2 Activity Diagram Pre Processing 32
3.3.1.4.3 Activity Diagram Manual Judgment 33

3.3.1.4.4 Activity Diagram Calculate Mutual Information 34
3.3.1.4.5 Activity Diagram Do Classification 35
3.3.2. Disain Penyimpanan Data 36
3.3.2.1. Entity Relationship Diagram (ERD) 36
3.3.2.2. Transfofmasi ERD ke dalam Tabel 36
3.3.2.2.1 Tabel Topic 36
3.3.2.2.2 Tabel Data Training 36
3.3.2.2.3 Tabel Data Testing 37
3.3.3. Disain Antarmuka 38
3.3.3.1. Rancangan Halaman Menu 38
3.3.3.2. Rancangan Halaman Get Training Data 38
3.3.3.3. Rancangan Halaman Manual Judgment 39
3.3.3.4. Rancangan Halaman Select Training Data 40
3.3.3.5. Rancangan Halaman Pre Processing 40
3.3.3.6. Rancangan Halaman Get Testing Data 41
3.3.3.7. Rancangan Halaman Analayze 42
3.3.3.8. Rancangan Halaman Streaming Tweets 42
BAB 4 PENGEMBANGAN PERANGKAT LUNAK 44
4.1. Implementasi Class 44
4.1.1. Class Pre Processing 45
4.1.1.1. Method Tokenizing 46
4.1.1.2. Method Formalization 46
4.1.1.3. Method Stopping 47
4.1.1.4. Method NGram 48
4.1.2. Class Mutual Information 49
4.1.3. Class Tweet 51
4.1.4. Class Attribute 52
4.1.5. Class Classification 52
4.1.6. Class Sentence 54
4.1.7. Class Gram 54
4.1.8. Class TestSentence 55
4.2. Implementasi Antarmuka 56

4.2.1. Halaman Menu 56
4.2.2. Halaman Get Training Data 57
4.2.3. Halaman Manual Judgment 58
4.2.4. Halaman Select Training Data 58
4.2.5. Halaman Pre Processing 59
4.2.6. Halaman Pre Processing Sentence 60
4.2.7. Halaman Pre Processing Tokenization 60
4.2.8. Halaman Pre Processing Formalization 61
4.2.9. Halaman Pre Processing Stopping 61
4.2.10. Halaman Pre Processing N Gram 62
4.2.11. Halaman Pre Processing Result MI 62
4.2.12. Halaman Pre Processing Selection 63
4.2.13. Halaman Get Testing Data 63
4.2.14. Halaman Analyze 64
4.2.15. Halaman Streaming Tweets 65
BAB 5 TESTING DAN EVALUASI SISTEM 66
5.1.1. Halaman Menu 66
5.1.2. Halaman Get Training Data 67
5.1.3. Halaman Manual Judgment 68
5.1.4. Halaman Select Training Data 69
5.1.5. Halaman Pre Processing 69
5.1.6. Halaman Get Testing Data 71
5.1.7. Halaman Analayze 72
5.1.8. Halaman Streaming Tweets 72
5.1.9. Hasil Manual Judgment Data Training 73
5.1.10. Hasil Testing Analisis Sentimen pada Data Training 76
5.1.11. Hasil Testing Analisis Sentimen pada Data Testing 76
5.1.12. Hasil Testing Analisis Sentimen pada Data Testing Stream 78
5.1.13. Hasil Testing Dengan Memperhitungkan Jumlah Keyword 79
5.2. Hasil Evaluasi Tingkat Keberhasilan 79
BAB 6 KESIMPULAN DAN SARAN 81
6.1. Kesimpulan 81

6.2. Saran..... 82
DAFTAR PUSTAKA 83

UNIVERSITAS KRISTEN MARANATHA

DAFTAR GAMBAR

Gambar 2.1 Contoh Autentikasi Twitter API 9
Gambar 2.2 Data Uji dengan Dua Kelas 10
Gambar 2.3 Dua Contoh Data Uji 11
Gambar 3.1 Arsitektur Pengambilan Data Training 24
Gambar 3.2 Arsitektur Pemilihan Term 25
Gambar 3.3 Arsitektur Proses Analisis Sentimen 27
Gambar 3.4 Use Case 28
Gambar 3.5 Activity Diagram Get Tweets 31
Gambar 3.6 Activity Diagram Pre Processing 32
Gambar 3.7 Activity Diagram Manual Judgment 33
Gambar 3.8 Activity Diagram Mutual Information 34
Gambar 3.9 Activity Diagram Do Classification 35
Gambar 3.10 Entity Relationship Diagram (ERD) 36
Gambar 3.11 Rancangan Halaman Menu 38
Gambar 3.12 Gambar Rancangan Get Training Data 38
Gambar 3.13 Rancangan Halaman Manual Judgment 39
Gambar 3.14 Rancangan Halaman Select Training Data 40
Gambar 3.15 Rancangan Halaman Pre Processing 40
Gambar 3.16 Rancangan Halaman Get Testing Data 41
Gambar 3.17 Rancangan Halaman Analyze 42
Gambar 3.18 Rancangan Halaman Streaming Tweet 42
Gambar 4.1 Class Diagram Sistem Analisis Sentimen Pada Twitter 44
Gambar 4.2 Class Preprocessing 45
Gambar 4.3 Method Tokenizing 46
Gambar 4.4 Method Formalization 47
Gambar 4.5 Method Stopping 48
Gambar 4.6 Method Unigram 48
Gambar 4.7 Method Bigram 49
Gambar 4.8 Class Mutual Information 49

Gambar 4.9 Method CalculateMI 50
Gambar 4.10 Class Tweet 51
Gambar 4.11 Class Attribute 52
Gambar 4.12 Class Classification 52
Gambar 4.13 Method testingNaiveBayes 53
Gambar 4.14 Class Sentence 54
Gambar 4.15 Class Gram 54
Gambar 4.16 Class TestSentence 55
Gambar 4.17 Halaman Menu 56
Gambar 4.18 Halaman Get Training Data 57
Gambar 4.19 Halaman Manual Judgment 58
Gambar 4.20 Halaman Select Training Data 58
Gambar 4.21 Halaman Pre Processing 59
Gambar 4.22 Halaman Pre Processing Sentence 60
Gambar 4.23 Halaman Pre Processing Tokenization 60
Gambar 4.24 Halaman Pre Processing Formalization 61
Gambar 4.25 Halaman Pre Processing Stopping 61
Gambar 4.26 Halaman Pre Processing N Gram 62
Gambar 4.27 Halaman Pre Processing Mutual Information 62
Gambar 4.28 Halaman Pre Processing Selection 63
Gambar 4.29 Halaman Get Data Testing 63
Gambar 4.30 Halaman Analyze 64
Gambar 4.31 Tampilan Streaming Tweets 65

DAFTAR TABEL

Tabel 2.1 Tabel Mutual Information 6
Tabel 3.1 Kumpulan Kalimat 13
Tabel 3.2 Kalimat setelah dilakukan Tokenizing 13
Tabel 3.3 Kalimat setelah proses Formalization 14
Tabel 3.4 Kalimat setelah proses Stopping 14
Tabel 3.5 Kalimat Setelah Proses Unigram dan Bigram 15
Tabel 3.6 Tabel *Unigram* Positif 16
Tabel 3.7 Tabel *Bigram* Positif 17
Tabel 3.8 Tabel *Unigram* Negatif 17
Tabel 3.9 Tabel *Bigram* Negatif 17
Tabel 3.10 Tabel *N-grams* Netral 18
Tabel 3.11 Tabel Hasil Mutual Information *Class* Positif 18
Tabel 3.12 Tabel Hasil Mutual Information *Class* Negatif 19
Tabel 3.13 Tabel Hasil Mutual Information *Class* Netral 19
Tabel 3.14 Tabel Data Training *Naïve Bayes* 19
Tabel 3.15 Tabel Proses Naïve Bayes 20
Tabel 3.16 Tabel Nilai Probabilistik 20
Tabel 3.17 Tabel Hasil Nilai Probabilistik dengan *Laplacian Smoothing* 21
Tabel 3.18 Tabel Data Testing dalam Format Naïve Bayes 21
Tabel 3.19 Tabel Topic 36
Tabel 3.20 Tabel Data Training 37
Tabel 3.21 Tabel Data Testing 37
Tabel 4.1 Fungsi Method dari Kelas Pre Processing 45
Tabel 4.2 Fungsi Method pada Kelas Mutual Information 50
Tabel 4.3 Fungsi Method pada Kelas Tweet 51
Tabel 5.1 Blackbox Testing Halaman Menu 66
Tabel 5.2 Blackbox Testing Halaman Get Training Data 67
Tabel 5.3 Blackbox Testing Halaman Manual Judgment 68
Tabel 5.4 Blackbox Testing Halaman Select Training Data 69

Tabel 5.5 Blackbox Testing Halaman Pre Processing 69
Tabel 5.6 Blackbox Testing Halaman Get Testing Data..... 71
Tabel 5.7 Blackbox Testing Halaman Analyze..... 72
Tabel 5.8 Blackbox Testing Halaman Streaming Tweets..... 72
Tabel 5.9 Contoh Retweet..... 73
Tabel 5.10 Contoh Tweets per Kategori 74
Tabel 5.11 Hasil Manual Judgment 75
Tabel 5.12 Hasil Persentase Manual Judgment..... 75
Tabel 5.13 Persentase Terkategori 76
Tabel 5.14 Matrix Confusion..... 76
Tabel 5.15 Hasil Testing pada Data Testing 77
Tabel 5.16 Hasil Testing yang Tidak Sesuai..... 77
Tabel 5.17 Hasil Streaming Tweets 78
Tabel 5.18 Hasil Persentasi dengan Memperhitungkan Jumlah Keyword 79
Tabel 5.19 Perbandingan Hasil Persentase Manual Judgment oleh Tiga Evaluator.. 80
Tabel 5.20 Hasil Testing Manual Judgment Evaluator A..... 80
Tabel 5.21 Hasil Testing Manual Judgment Evaluator B 80
Tabel 5.22 Hasil Testing Manual Judgment Evaluator C 80

DAFTAR NOTASI/LAMBANG

Jenis	Notasi/Lambang	Nama	Arti
ERD		Entitas	Menunjukkan sebuah objek yang dapat dibedakan dengan objek lainnya
ERD		Atribut	Mendeskripsikan karakter entitas
ERD		Relasi	Menunjukkan adanya hubungan diantara sejumlah entitas yang berbeda
ERD		Garis (*one to many*)	Penghubung antar relasi dan entitas dimana satu entitas dapat memiliki lebih dari satu hubungan.