

**DETEKSI EMOSI DARI *TWEET* BERBAHASA INDONESIA
MENGUNAKAN LDA DAN KONVERSI *EXPRESSION SYMBOL***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
Pada Departemen Ilmu Komputer/ Informatika**

Disusun oleh :

RISMA MUSTIKA CAHYANINGTYAS

24010312130072

DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

2017

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Risma Mustika Cahyaningtyas

NIM : 24010312130072

Judul : Deteksi Emosi dari *Tweet* Berbahasa Indonesia Menggunakan LDA dan Konversi *Expression Symbol*

Dengan ini saya menyatakan bahwa dalam tugas akhir/ skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.



HALAMAN PENGESAHAN

Judul : Deteksi Emosi dari *Tweet* Berbahasa Indonesia Menggunakan LDA dan Konversi

Expression Symbol

Nama : Risma Mustika Cahyaningtyas

NIM : 24010312130072

Telah diujikan pada sidang tugas akhir tanggal 10 Juli 2017 dan dinyatakan lulus pada tanggal 10 Juli 2017.

Semarang, 26 Juli 2017



Panitia Penguji Tugas Akhir

Ketua,

Suknawati Nur Endah, S.Si, M.Kom
NIP. 197805022005012002

HALAMAN PENGESAHAN

Judul : Deteksi Emosi dari *Tweet* Berbahasa Indonesia Menggunakan LDA dan Konversi
Expression Symbol

Nama : Risma Mustika Cahyaningtyas

NIM : 24010312130072

Telah diujikan pada sidang tugas akhir tanggal 10 Juli 2017.

Semarang, 26 Juli 2017

Pembimbing,



Dr. Retno Kusumaningrum, S.Si, M.Kom
NIP. 198104202005012001

ABSTRAK

Twitter sebagai salah satu jejaring sosial yang menarik perhatian banyak masyarakat Indonesia karena dianggap sebagai tempat berbagi karya, ide, opini tentang isu-isu yang terjadi, dan media untuk mengungkapkan berbagai hal mengenai kehidupan pribadi. Kepopuleran *Twitter* dapat digunakan menjadi sumber data pendapat dan sentimen masyarakat yang efisien untuk pemasaran ataupun studi sosial. Salah satu bentuk studi sosial yang dapat diterapkan pada proses analisis *twitter* adalah deteksi emosi. Pendeteksian emosi berpotensi untuk diterapkan dalam berbagai macam aplikasi mulai dari aplikasi kesehatan, konseling, bisnis, hingga studi populasi masyarakat. Penelitian ini memanfaatkan salah satu model pemodelan topik terpopuler dan paling sederhana yaitu *Latent Dirichlet Allocation* (LDA) serta konversi *expression symbol* (*emoticon/ emoji*) yang menunjukkan emosi ataupun topik pada sebuah *tweet* untuk memperbanyak kosa kata yang merepresentasikan emosi. Kelebihan dari metode LDA yang diajukan adalah dapat mendeteksi beberapa emosi pada *tweet* karena pendeteksian yang dilakukan tidak bersifat kaku dan dapat menunjukkan proporsi emosi yang ada pada *tweet*. Penelitian ini juga membandingkan deteksi emosi menggunakan LDA dan konversi *expression symbol* dengan deteksi emosi menggunakan LDA (tanpa konversi *expression symbol*). Hasil penelitian menunjukkan bahwa deteksi emosi menggunakan LDA dan konversi *expression symbol* lebih baik dengan rata-rata selisih akurasi mencapai sebesar 14.096%.

Kata kunci : *Latent Dirichlet Allocation*, deteksi emosi, *twitter*, konversi *expression symbol*

ABSTRACT

Twitter as one of the social networks that attracts many Indonesian people because it is considered as a place to share works, ideas, opinions about the issues that occur, and the media to express various things about personal life. Twitter's popularity can be used as an efficient source of people's opinion and sentiment for marketing or social studies. One form of social studies that can be applied to the process of twitter analysis is emotional detection. Emotional detection has a potency to be applied in a wide range of applications ranging from health applications, counseling, business, to community population studies. This research utilizes one of the most popular and simplest topic modeling models of Latent Dirichlet Allocation (LDA) method and conversion expression symbol (emoticon / emoji) which shows the emotion or topic in a tweet to multiply the vocabulary that represents emotion. The advantages of the LDA method proposed is that it can detect some emotion on the tweet because the detection is not rigid and is able to show the proportion of emotion in the tweet. This research also compares emotional detection using LDA and conversion expression symbol with emotional detection using LDA (without conversion expression symbol). The results show that emotional detection using LDA and conversion expression symbol is better with the average difference of accuracy reached 14.096%.

Keywords: Latent Dirichlet Allocation, emotion detection, twitter, conversion expression symbol

KATA PENGANTAR

Keanugrahan inspirasi dari Tuhan Yang Maha Agung menjadi kekuatan kepada penulis untuk menyelesaikan laporan tugas akhir yang berjudul “Deteksi Emosi dari *Tweet* Berbahasa Indonesia Menggunakan LDA dan Konversi *Expression Symbol*”. Laporan tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan laporan ini penulis banyak mendapat bimbingan dan bantuan dari berbagai pihak. Untuk itu, pada kesempatan ini penulis mengucapkan rasa hormat dan terima kasih kepada :

1. Ragil Saputra, S.Si, M.Cs selaku ketua Departemen Ilmu Komputer/ Informatika
2. Helmie Arif Wibawa, S.Si, M.Cs selaku Koordinator Tugas Akhir
3. Dr. Retno Kusumaningrum, S.Si, M.Kom selaku dosen pembimbing yang telah memberikan waktu dan bimbingan yang tak ternilai bagi penulis
4. Reza Krisnadi Pratama yang telah membantu dalam penentuan emosi sebagai bagian dalam proses pengumpulan data
5. Semua pihak yang telah membantu kelancaran dalam penyusunan tugas akhir, yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam laporan ini masih banyak kekurangan baik dari segi materi ataupun dalam penyajiannya karena keterbatasan kemampuan dan pengetahuan penulis. Oleh karena itu, kritik dan saran sangat penulis harapkan. Semoga laporan ini dapat bermanfaat bagi pembaca dan penulis pada umumnya.

Semarang, 26 Juli 2017

Penulis,

Risma Mustika Cahyaningtyas

24010312130072

DAFTAR ISI

| | |
|--|------|
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI..... | ii |
| HALAMAN PENGESAHAN | iii |
| HALAMAN PENGESAHAN | iv |
| ABSTRAK | v |
| ABSTRACT | vi |
| KATA PENGANTAR..... | vii |
| DAFTAR ISI | viii |
| DAFTAR GAMBAR..... | xi |
| DAFTAR TABEL | xiii |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah..... | 4 |
| 1.3 Tujuan dan Manfaat | 4 |
| 1.4 Ruang Lingkup | 4 |
| 1.5 Sistematika Penulisan | 4 |
| BAB II TINJAUAN PUSTAKA | 6 |
| 2.1 Perkembangan Deteksi Emosi pada <i>Twitter</i> | 6 |
| 2.2 Emosi | 7 |
| 2.3 <i>Twitter</i> | 8 |
| 2.4 <i>Latent Dirichlet Allocation</i> | 8 |
| 2.5 <i>Stemming</i> | 10 |
| 2.6 Evaluasi..... | 11 |
| 2.7 Pengembangan Perangkat Lunak..... | 12 |
| 2.7.1 Fase RUP | 13 |
| 2.7.2 Proses Iteratif RUP | 15 |
| 2.8 Pemrograman Berorientasi Objek..... | 16 |
| 2.9 <i>Unified Modeling Language (UML)</i> | 18 |
| 2.9.1 <i>Use Case Diagram</i> | 18 |
| 2.9.2 <i>Activity Diagram</i> | 19 |
| 2.9.3 <i>Class Diagram</i> | 20 |
| 2.9.4 <i>Sequence Diagram</i> | 21 |
| BAB III METODOLOGI PENELITIAN | 23 |

| | | |
|---------------------------------|--|----|
| 3.1 | Proses Pembentukan Model Deteksi..... | 23 |
| 3.1.1 | Pengumpulan Data..... | 23 |
| 3.1.2 | Pembagian Data..... | 24 |
| 3.1.3 | <i>Preprocessing</i> | 25 |
| 3.1.3.1 | Konversi <i>Expression Symbol</i> | 25 |
| 3.1.3.2 | Konversi <i>Hashtag</i> | 27 |
| 3.1.3.3 | Tokenisasi..... | 29 |
| 3.1.3.4 | Penghilangan <i>Stopword</i> | 30 |
| 3.1.3.5 | <i>Stemming</i> | 31 |
| 3.1.4 | Pelatihan | 33 |
| 3.1.5 | Pengujian dan Evaluasi..... | 38 |
| 3.2 | Deteksi Emosi | 43 |
| 3.3 | Fase <i>Inception</i> | 44 |
| 3.3.1 | <i>Business Modeling</i> | 44 |
| 3.3.2 | Definisi Kebutuhan..... | 44 |
| 3.3.2.1 | Deskripsi Umum | 45 |
| 3.3.2.2 | Kebutuhan Fungsional Sistem..... | 45 |
| 3.3.2.3 | Kebutuhan Non Fungsional Sistem..... | 45 |
| 3.3.2.4 | Model <i>Use Case</i> | 45 |
| 3.4 | Fase <i>Elaboration</i> | 47 |
| 3.4.1 | Tahap Analisis | 47 |
| 3.4.1.1 | Realisasi <i>Use Case</i> Tahap Analisis | 47 |
| 3.4.1.2 | <i>Class Analisis</i> | 48 |
| 3.4.1.3 | <i>Sequence Diagram</i> | 50 |
| 3.4.1.4 | <i>Activity Diagram</i> | 50 |
| 3.4.2 | Tahap Desain | 50 |
| 3.4.2.1 | Realisasi <i>Use Case</i> Tahap Desain | 50 |
| 3.4.2.2 | <i>Class</i> Desain..... | 51 |
| 3.4.2.3 | <i>Class Diagram</i> | 52 |
| 3.4.2.4 | Perancangan Antarmuka | 52 |
| 3.4.2.5 | Pemodelan Data..... | 56 |
| BAB IV HASIL DAN ANALISIS | | 58 |
| 4.1 | Fase <i>Construction</i> | 58 |
| 4.1.1 | Lingkungan Implementasi | 58 |

| | | |
|--|--|-----|
| 4.1.2 | Implementasi Data | 58 |
| 4.1.3 | Implementasi <i>Class</i> | 59 |
| 4.1.4 | Implementasi Antarmuka | 60 |
| 4.2 | Skenario Pengujian | 64 |
| 4.2.1 | Pengujian Fungsional Sistem | 64 |
| 4.2.2 | Pengujian Kinerja Sistem | 65 |
| 4.2.2.1 | Data | 65 |
| 4.2.2.2 | Eksperimen..... | 65 |
| 4.3 | Hasil dan Analisis Sistem | 66 |
| 4.3.1 | Pengujian Fungsional Sistem | 66 |
| 4.3.2 | Hasil dan Analisis untuk Skenario Eksperimen 1 | 68 |
| 4.3.3 | Hasil dan Analisis untuk Skenario Eksperimen 2 | 70 |
| 4.3.4 | Hasil dan Analisis untuk Skenario Eksperimen 3 | 72 |
| BAB V PENUTUP | | 77 |
| 5.1 | Kesimpulan | 77 |
| 5.2 | Saran | 77 |
| DAFTAR PUSTAKA..... | | 78 |
| LAMPIRAN-LAMPIRAN | | 80 |
| Lampiran 1. Data Deteksi Emosi..... | | 81 |
| Lampiran 2. Konversi <i>Expression Symbol</i> | | 88 |
| Lampiran 3. <i>Hashtag</i> | | 91 |
| Lampiran 4. <i>Stemming</i> | | 95 |
| Lampiran 5. Contoh Perhitungan Proses Deteksi | | 97 |
| Lampiran 6. Fase <i>Elaboration</i> | | 117 |
| Lampiran 7. Sistem Pembentukan Model Deteksi Emosi | | 121 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1. Perbedaan LDA sebagai Proses Generatif dan Inferensi (Steyvers & Griffiths, 2007) | 8 |
| Gambar 2.2. 3-fold Cross Validation..... | 12 |
| Gambar 2.3. Alur Hidup RUP (Sommerville, 2011)..... | 13 |
| Gambar 2.4. Proses Iteratif RUP (Sukanto & Shalahuddin, 2013) | 15 |
| Gambar 3.1. Tahapan Pembentukan Model Deteksi Emosi dan Proses Deteksi Emosi | 23 |
| Gambar 3.2. Flowchart Preprocessing..... | 25 |
| Gambar 3.3. <i>Flowchart</i> Proses Konversi <i>Expression Symbol</i> | 26 |
| Gambar 3.4. <i>Flowchart</i> Ubah Simbol | 27 |
| Gambar 3.5. <i>Flowchart</i> Proses Konversi <i>Hashtag</i> | 28 |
| Gambar 3.6. <i>Flowchart</i> Proses Tokenisasi..... | 29 |
| Gambar 3.7. <i>Flowchart</i> Proses Penghilangan <i>Stopword</i> | 30 |
| Gambar 3.8. <i>Flowchart</i> Hapus <i>Stopword</i> | 31 |
| Gambar 3.9. <i>Flowchart</i> Hapus Kata Ganti..... | 31 |
| Gambar 3.10. <i>Flowchart</i> Proses <i>Stemming</i> | 32 |
| Gambar 3.11. <i>Flowchart</i> Proses Pelatihan | 33 |
| Gambar 3.12. Proses <i>Gibbs Sampling</i> | 34 |
| Gambar 3.13. <i>Flowchart</i> Proses Inisialisasi Topik | 34 |
| Gambar 3.14. <i>Flowchart</i> Hitung <i>Doctopic</i> | 35 |
| Gambar 3.15. <i>Flowchart</i> Hitung <i>Topicperword</i> | 35 |
| Gambar 3.16. <i>Flowchart</i> Hitung <i>Wordsintopic</i> | 36 |
| Gambar 3.17. <i>Flowchart</i> Normalisasi Probabilitas Topik..... | 36 |
| Gambar 3.18. <i>Flowchart</i> Memilih Topik Baru | 37 |
| Gambar 3.19. <i>Flowchart</i> Hitung Probabilitas Kata Terhadap Topik (<i>PWZ</i>), Probabilitas Topik Terhadap Dokumen (<i>PZD</i>), dan Probabilitas Setiap Topik (<i>PZ</i>) | 37 |
| Gambar 3.20. <i>Flowchart</i> Hitung <i>PZC</i> (Probabilitas Topik Terhadap Setiap Kelas)..... | 38 |
| Gambar 3.21. <i>Flowchart</i> Tahap Pengujian dan Evaluasi | 39 |
| Gambar 3.22. <i>Flowchart</i> Cari <i>PWZ</i> (Probabilitas Kata Terhadap Topik)..... | 39 |
| Gambar 3.23. <i>Flowchart</i> Hitung <i>PZD</i> (Probabilitas Topik Terhadap Dokumen)..... | 40 |
| Gambar 3.24. <i>Flowchart</i> Hitung Similaritas Menggunakan <i>KLD</i> | 41 |
| Gambar 3.25. <i>Flowchart</i> Menentukan Topik (Kelas) | 41 |
| Gambar 3.26. <i>Flowchart</i> Evaluasi..... | 42 |

| | |
|---|----|
| Gambar 3.27. Flowchart Bentuk <i>Confussion Matrix</i> | 42 |
| Gambar 3.28. <i>Flowchart</i> Hitung Akurasi..... | 42 |
| Gambar 3.29. <i>Flowchart</i> Deteksi Emosi | 43 |
| Gambar 3.30. Business Use Case Diagram | 44 |
| Gambar 3.31. Diagram <i>Use Case</i> | 46 |
| Gambar 3.32. <i>Analysis Class Diagram</i> Deteksi Emosi..... | 47 |
| Gambar 3.33. <i>Analysis Class Diagram</i> Menampilkan Hasil..... | 48 |
| Gambar 3.34. Desain Antarmuka Halaman Utama | 52 |
| Gambar 3.35. Desain Antarmuka Halaman Deteksi..... | 53 |
| Gambar 3.36. Desain Antarmuka Hasil Deteksi..... | 53 |
| Gambar 3.37. Desain Halaman Data Deteksi | 54 |
| Gambar 3.38. Desain Halaman Praproses Deteksi | 54 |
| Gambar 3.39. Desain Halaman Detail Nilai PWZ..... | 55 |
| Gambar 3.40. Desain Halaman Detail Hasil Nilai PZD | 55 |
| Gambar 3.41. Desain Halaman Detail Hasil Deteksi. | 56 |
| Gambar 4.1. Implementasi Tabel Kata Dasar..... | 58 |
| Gambar 4.2. Implementasi Tabel Kata..... | 59 |
| Gambar 4.3. Implementasi Tabel <i>Stoplist</i> | 59 |
| Gambar 4.4. Implementasi Tabel Kata Ganti | 59 |
| Gambar 4.5. Implmentasi Halaman Utama | 60 |
| Gambar 4.6. Implementasi Halaman Deteksi..... | 61 |
| Gambar 4.7. Implementasi Halaman Hasil Deteksi | 61 |
| Gambar 4.8. Implementasi Halaman Data Deteksi | 62 |
| Gambar 4.9. Implementasi Halaman Praproses Deteksi | 62 |
| Gambar 4.10. Implementasi Halaman Detail Nilai PWZ..... | 63 |
| Gambar 4.11. Implementasi Halaman Detail Hasil Nilai PZD | 63 |
| Gambar 4.12. Implementasi Halaman Detail Hasil Deteksi..... | 64 |
| Gambar 4.13. Skenario Eksperimen | 65 |
| Gambar 4.14. Grafik Hasil Nilai Probabilitas Kelas Terhadap Topik (<i>PZC</i>) Skenario 1 ... | 69 |
| Gambar 4.15. Grafik Hasil Nilai Probabilitas Kelas Terhadap Topik (<i>PZC</i>) Skenario 2 ... | 71 |
| Gambar 4.16. Grafik Perbandingan Selisish Nilai Akurasi..... | 73 |
| Gambar 4.17. Grafik Perbandingan Selisih Nilai Akurasi | 74 |
| Gambar 4.18. Grafik Perbandingan Selisih Nilai Akurasi | 75 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1. Penelitian Deteksi Emosi..... | 6 |
| Tabel 2.2. Ekspresi Wajah dari Emosi Dasar Manusia | 7 |
| Tabel 2.3. Keterangan Notasi Persamaan 2.1, 2.2 dan 2.3..... | 10 |
| Tabel 2.4. <i>Confussion Matrix</i> | 11 |
| Tabel 2.5. Keterangan Notasi pada Tabel 2.4 dan Persamaan 2.4 | 12 |
| Tabel 2.6. Keterangan Proses Iteratif RUP..... | 16 |
| Tabel 2.7. Simbol yang Digunakan dalam Diagram <i>Use Case</i> (Sukamto & Shalahuddin, 2013)..... | 19 |
| Tabel 2.8. Format Skenario <i>Use Case</i> (Sukamto & Shalahuddin, 2013) | 19 |
| Tabel 2.9. Simbol-simbol pada Diagram Aktivitas (Sukamto & Shalahuddin, 2013) | 20 |
| Tabel 2.10. Simbol-simbol pada Diagram Kelas. (Sukamto & Shalahuddin, 2013)..... | 20 |
| Tabel 2.11. Simbol-simbol pada Diagram Sekuen (Sukamto & Shalahuddin, 2013)..... | 21 |
| Tabel 3.1. Pembagian Data Latih | 24 |
| Tabel 3.2. Pembagian Data Uji..... | 24 |
| Tabel 3.3. Contoh Konversi <i>Expression Symbol</i> | 27 |
| Tabel 3.4. Contoh Hasil Konversi Hashtag | 28 |
| Tabel 3.5. Contoh Hasil Proses Tokenisasi | 30 |
| Tabel 3.6. Contoh Hasil Penghilangan <i>Stopword</i> | 31 |
| Tabel 3.7. Contoh Hasil Proses <i>Stemming</i> | 32 |
| Tabel 3.8. Kombinasi Input | 33 |
| Tabel 3.9. Daftar Aktor..... | 46 |
| Tabel 3.10. Daftar <i>Use Case</i> | 46 |
| Tabel 3.11. Detail <i>Use Case</i> Deteksi Emosi..... | 46 |
| Tabel 3.12. Detail <i>Use Case</i> Menampilkan Hasil | 47 |
| Tabel 3.13. Identifikasi <i>Class Analisis</i> | 48 |
| Tabel 3.14. Tanggungjawab dan Atribut <i>Class</i> | 49 |
| Tabel 3.15. Identifikasi <i>Class</i> Desain Deteksi Emosi | 51 |
| Tabel 3.16. Identifikasi <i>Class</i> Desain Menampilkan Hasil | 51 |
| Tabel 3.17. Daftar <i>Class</i> Desain..... | 51 |
| Tabel 3.18. Struktur Tabel <i>tb_stoplist</i> | 56 |
| Tabel 3.19. Struktur Tabel <i>tb_kataganti</i> | 57 |
| Tabel 3.20. Struktur Tabel <i>tb_kata</i> | 57 |

| | |
|--|----|
| Tabel 3.21. Struktur Tabel <i>tb_katadasar</i> | 57 |
| Tabel 4.1. Implementasi <i>class</i> | 59 |
| Tabel 4.2. Rencana Pengujian Fungsional Aplikasi | 64 |
| Tabel 4.3. Hasil Pengujian Fungsionalitas | 67 |
| Tabel 4.4. Hasil Nilai Akurasi Skenario 1 | 68 |
| Tabel 4.5. Hasil Nilai <i>PWZ</i> (Probabilitas Kata Pada Setiap Topik) Kombinasi 1 <i>Fold</i> 1... | 69 |
| Tabel 4.6. Tabel Hasil Nilai Akurasi Skenario 2..... | 70 |
| Tabel 4.7. Hasil Nilai <i>PWZ</i> (Probabilitas Kata Pada Setiap Topik) Kombinasi 1 <i>Fold</i> 4... | 71 |
| Tabel 4.8. Tabel Selisih Nilai Akurasi Skenario 1 dan 2 | 72 |

BAB I

PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup serta sistematika penulisan dari penelitian tugas akhir mengenai pendeteksian emosi dari *tweet* berbahasa Indonesia menggunakan *Latent Dirichlet Allocation* (LDA) dan konversi *expression symbol*.

1.1 Latar Belakang

Seiring dengan kemajuan teknologi yang cepat, pengguna internet di Indonesia pun semakin meningkat. Hasil riset dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menyatakan bahwa pengguna internet hingga akhir tahun 2016 mengalami pertumbuhan 44,6 juta pengguna, yaitu dari 88,1 juta menjadi 132,7 juta pengguna. Sebesar 97,4% dari pengguna internet di Indonesia mengakses aplikasi/konten jejaring sosial (Asosiasi Penyelenggara Jasa Internet Indonesia, 2016). Jejaring sosial banyak menyita perhatian masyarakat Indonesia karena dianggap dapat menjadi wadah bagi karya, ide, opini tentang isu-isu yang terjadi secara bebas, dan media untuk mengungkapkan berbagai hal mengenai kehidupan pribadi mereka.

Salah satu jejaring sosial yang banyak digunakan masyarakat Indonesia adalah *Twitter*. CEO *Twitter* Dick Costolo mengungkapkan bahwa saat ini jumlah pengguna *Twitter* di Indonesia sudah mencapai 50 juta anggota. *Twitter* memungkinkan penggunaannya untuk berbagi pesan menggunakan teks sepanjang 140 karakter yang disebut *tweet*. Dalam sebuah *tweet*, pengguna biasanya juga menggunakan *emoticon/emoji* dan *hashtag* untuk menunjukkan emosi ataupun topik dari *tweet* mereka.

Twitter dapat digunakan menjadi sumber data pendapat dan sentimen masyarakat yang efisien untuk pemasaran atau studi sosial karena merupakan sarana yang digunakan banyak orang untuk mengekspresikan opini, memiliki data teks yang diunggah oleh penggunaannya dalam skala besar dan bertambah setiap harinya, dan penggunaannya pun berasal dari berbagai kalangan (Pak & Paroubek, 2010).

Salah satu bentuk studi sosial yang dapat diterapkan pada proses analisis twitter adalah deteksi emosi. Emosi merupakan keadaan kompleks dari pikiran yang dipengaruhi oleh peristiwa eksternal, perubahan fisiologis, atau hubungan dengan orang lain. Ekman (1999) mengklasifikasikan emosi manusia pada enam kategori,

yaitu *happiness* (kebahagiaan), *anger* (kemarahan), *sadness* (kesedihan), *fear* (ketakutan), *disgust* (jijik), dan *surprise* (terkejut).

Pendeteksian emosi berpotensi untuk diterapkan dalam berbagai macam aplikasi mulai dari aplikasi kesehatan, konselor, sampai studi populasi masyarakat. Pendeteksian emosi dapat digunakan oleh profesional di bidang kesehatan atau lembaga konseling untuk memonitor dan melacak kondisi emosional pasien atau untuk mengenali kecemasan atau stres sistemik dari penduduk (misalnya kelompok-kelompok mahasiswa yang berbeda di kampus). Sistem ini juga dapat membantu lembaga komersial untuk mengukur sentimen pembeli atau untuk memfasilitasi produk iklan yang ditargetkan. Selain itu, teknologi ini dapat mengukur suasana hati dari orang-orang dalam suatu komunitas, yang mungkin membantu ilmuwan sosial untuk memahami kualitas hidup dari penduduk. Mengukur dan melacak kondisi dan kualitas hidup dapat dilakukan berdasarkan berbagai aspek kehidupan seperti sosial, emosional, psikologi, kepuasan hidup, dan pekerjaan. Namun, metode yang mengukur kondisi hidup gagal untuk mengukur apa yang orang pikirkan dan rasakan tentang hidup mereka, seperti emosi positif atau negatif, atau kepuasan akan hidup mereka. Kualitas hidup biasanya diukur menggunakan survei. Masyarakat diminta mengisi kuesioner tentang kehidupan dan emosi mereka sehari-hari. Mengumpulkan kuesioner ini sangat memakan waktu dan rawan kesalahan. Pendeteksian emosi pada jejaring sosial seperti twitter dapat menjadi solusi untuk hal ini. Hal tersebut dikarenakan tujuan dan isi sebuah *tweet* dapat mendiskripsikan emosi penggunanya (Hasan, et al., 2014).

Beberapa peneliti telah meneliti twitter untuk menyelidiki potensi penggunaan media sosial untuk mendeteksi gangguan depresi. Park, Cha, dan Cha (2012) melakukan beberapa studi untuk menangkap suasana depresi dari pengguna *Twitter*, dengan memahami *personal updates* dari beberapa individu yang menjadi objek penelitian menggunakan pendekatan psikologi. Hasil penelitian tersebut menunjukkan bahwa individu mengalami depresi meningkatkan penggunaan kata yang berhubungan dengan emosi negatif dan kemarahan pada *tweet* mereka (Park, et al., 2012). Sedangkan Purver dan Battersby (2012) menerapkan metode *supervised classifier* untuk deteksi emosi pada pesan *Twitter* menggunakan data berlabel otomatis. Penelitian tersebut menggunakan enam emosi dasar yang diusulkan oleh Ekman (1972). Metode tersebut memiliki kinerja yang baik (akurasi mencapai 60%) untuk

beberapa emosi seperti kebahagiaan, kesedihan, dan kemarahan dibandingkan takut, terkejut, dan jijik (Purver & Battersby, 2012). Hasan, Rundensteiner, dan Agu (2014) menerapkan *hashtag* pada *Twitter* untuk memberi label otomatis pada pesan sesuai kelas emosional dan melatih pengklasifikasi untuk deteksi emosi multikelas. Hasil penelitian mereka menunjukkan bahwa *hashtag* dan penanda konvensional lainnya dari *tweets* adalah fitur yang berguna untuk sentimen dan klasifikasi emosi (Hasan, et al., 2014). Arifin, Sari, dan Ratnasari (2014) mengelompokkan *tweet* berbahasa Indonesia dalam lima emosi, seperti senang, marah, takut, sedih, dan terkejut. Penelitian tersebut menilai pendeteksian emosi menggunakan makna dari *tweet* saja masih kurang, sehingga mendapatkan fitur yang sesuai dan berkaitan dengan emosi (seperti *hashtag*, *emoji*, *emoticon*, dan kata sifat) merupakan hal yang sangat penting pada fase ekstraksi pertama. Penelitian tersebut juga menggunakan *Non-Negative Matrix Factorization* untuk menganalisa keterkaitan antar fitur dan mereduksi fitur secara semantik yang kemudian diklasifikasikan menggunakan *K-Nearest Neighbor* (KNN) (Arifin, et al., 2014).

Berdasarkan berbagai penelitian tersebut, fitur yang sesuai sangat diperlukan untuk melakukan proses pendeteksian emosi. Penelitian tersebut juga hanya mendeteksi satu emosi pada setiap *tweet*, karena proses deteksi yang dilakukan bersifat kaku. Namun, pada umumnya pengguna *twitter* dapat mengungkapkan berbagai emosi mereka dalam sebuah *tweet*. Sehingga jika melakukan deteksi yang kaku, maka emosi lain yang terkandung di dalam sebuah *tweet* tidak dapat diketahui.

Probabilistic topics modeling telah menghasilkan efek yang cukup baik pada berbagai aplikasi yang ada khususnya dalam bidang pengklasifikasian teks dan *Information Retrieval*. Salah satu model dari pemodelan topik yaitu *Latent Dirichlet Allocation* yang merupakan pemodelan topik terpopuler dan paling sederhana. Selain itu, LDA memungkinkan dokumen memiliki beberapa topik dengan proporsi yang berbeda dan juga mengatasi *overfitting* yang disebabkan oleh pertumbuhan parameter secara linear (Blei, 2012). LDA merupakan model umum probabilitas untuk sekumpulan data diskret seperti kumpulan dokumen. Dalam model probabilitas tersebut variabel laten (tersembunyi) dapat menjelaskan variabel observasi (teramati). Variabel observasi adalah dokumen, sementara topik yang ditentukan dari tiap kata pada dokumen adalah variabel laten. Variabel laten tersebut adalah parameter yang harus diestimasi dari model LDA yang didapatkan.

Oleh karena itu, penelitian yang akan dilakukan adalah mendeteksi emosi pada jejaring sosial *Twitter* menggunakan metode LDA dan memanfaatkan konversi *expression symbol* (*emoji/ emoticon*) untuk memperbanyak kosa kata yang merepresentasikan emosi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan maka dapat dibuat rumusan masalah yaitu bagaimana mendeteksi emosi dari *tweet* berbahasa Indonesia menggunakan konversi *expression symbol* dan *Latent Dirichlet Allocation*.

1.3 Tujuan dan Manfaat

Tujuan dari penelitian tugas akhir ini adalah sebagai berikut :

1. Mengetahui kinerja deteksi emosi dari *tweet* berbahasa Indonesia dengan menerapkan metode LDA dan konversi *expression symbol*.
2. Mengetahui kinerja deteksi emosi dari *tweet* berbahasa Indonesia dengan metode LDA tanpa menerapkan konversi *expression symbol*.
3. Membandingkan kinerja penerapan konversi *expression symbol* pada deteksi emosi dari *tweet* berbahasa Indonesia.

Adapun manfaat dari penelitian tugas akhir ini adalah untuk mengidentifikasi profil emosi dari pengguna *Twitter* untuk kurun waktu satu tahun terakhir.

1.4 Ruang Lingkup

Ruang lingkup dari penelitian Tugas Akhir ini adalah sebagai berikut :

1. *Tweet* yang akan dideteksi hanyalah *tweet* berbahasa Indonesia dan **bukan** merupakan *tweet* yang hanya mengandung satu kata, *link*, *retweet*, dan kata ganti orang informal.
2. Emosi yang akan dideteksi hanyalah emosi senang, marah, takut, sedih, dan terkejut.

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi dalam beberapa pokok bahasan, yaitu:

- BAB I PENDAHULUAN**
- Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup serta sistematika penulisan dalam pembuatan tugas akhir mengenai deteksi emosi dari *tweet* berbahasa Indonesia menggunakan LDA dan konversi *expression symbol*.
- BAB II TINJAUAN PUSTAKA**
- Bab ini menyajikan tinjauan pustaka yang berhubungan dengan topik tugas akhir. Tinjauan pustaka yang digunakan dalam penyusunan tugas akhir ini meliputi penjelasan mengenai penelitian yang terkait dengan deteksi emosi pada *Twitter*, emosi, *Twitter*, *Latent Dirichlet Allocation*, evaluasi, pengembangan perangkat lunak, pemograman berorientasi objek dan *Unified Modeling Language* (UML).
- BAB III METODOLOGI PENELITIAN**
- Bab ini menyajikan tahapan yang dilakukan pada penelitian tugas akhir. Pada bab ini akan dijelaskan mengenai gambaran umum penelitian dan tahap-tahap yang dilakukan selama melakukan penelitian seperti pengumpulan data, preprocessing, pembentukan model deteksi emosi, serta proses deteksi.
- BAB IV HASIL DAN ANALISIS**
- Bab ini menyajikan pembahasan mengenai data, skenario, hasil dan analisis penelitian serta pengembangan perangkat lunak yang dilakukan dalam penelitian tugas akhir.
- BAB V PENUTUP**
- Bab ini merupakan kesimpulan dari bab-bab yang dibahas sebelumnya dan saran untuk pengembangan penelitian lebih lanjut.