

**ANALISIS SENTIMEN MENGGUNAKAN *LATENT DIRICHLET*
ALLOCATION DAN VISUALISASI *TOPIC POLARITY WORDCLOUD***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer/ Informatika**

Disusun oleh :

MOHAMMAD FAJAR AINUL BASHRI

24010312120002

DEPARTEMEN ILMU KOMPUTER/INFORMATIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

2017

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Mohammad Fajar Ainul Bashri

NIM : 24010312120002

Judul : Analisis Sentimen Menggunakan *Latent Dirichlet Allocation* dan Visualisasi
Topic Polarity Wordcloud

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 dalam daftar pustaka.

Semarang, 3 Februari 2017



Mohammad Fajar Ainul Bashri

NIM 24010312120002

HALAMAN PENGESAHAN

Judul : Analisis Sentimen Menggunakan *Latent Dirichlet Allocation* dan Visualisasi
Topic Polarity Wordcloud

Nama : Mohammad Fajar Ainul Bashri

NIM : 24010312120002

Telah diujikan pada sidang tugas akhir pada tanggal 19 Januari 2017 dan dinyatakan lulus pada tanggal 19 Januari 2017.

Semarang, 3 Februari 2017

Mengetahui,

Ketua Departemen Ilmu Komputer/Informatika



Ragil Saputra, S.Si., M.Cs

NIP. 19801021 200501 1003

Ketua Penguji Tugas Akhir

A handwritten signature in blue ink, which appears to read 'Wibawa', is written over the text.

Helmie Arif Wibawa, S.Si., M.Cs

NIP. 19780516 200312 1001

HALAMAN PENGESAHAN

Judul : Analisis Sentimen Menggunakan *Latent Dirichlet Allocation* dan Visualisasi
Topic Polarity Wordcloud

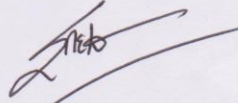
Nama : Mohammad Fajar Ainul Bashri

NIM : 24010312120002

Telah diujikan pada sidang tugas akhir pada tanggal 19 Januari 2017.

Semarang, 3 Februari 2017

Dosen Pembimbing,



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

NIP. 19810420 200501 2001

ABSTRAK

Bagi sebuah institusi pendidikan, pendapat peserta didik memiliki peran penting untuk kemajuan institusi pendidikan tersebut. Analisis sentimen dapat membantu dalam mengekstraksi pendapat yang berasal dari sebuah dokumen, komentar, *review* produk, dan data-data lainnya sehingga akan lebih mudah dalam melakukan pemantauan. Salah satu metode untuk melakukan analisis sentimen adalah *Latent Dirichlet Allocation* (LDA) yang mampu digunakan untuk mengekstraksi topik dari kumpulan dokumen komentar, dimana topik tersebut direpresentasikan sebagai kemunculan kata-kata dengan probabilitas topik yang berbeda. Oleh karena itu diperlukan representasi data dalam bentuk visual yang mudah dipahami daripada teks dan tabel. Salah satu bentuk visualisasi data adalah *wordcloud* yang memberikan gambaran visual frekuensi kemunculan kata. Pada penelitian ini akan dilakukan analisis sentimen terhadap komentar mahasiswa pada suatu universitas, dalam hal ini Universitas Diponegoro, menggunakan LDA dengan visualisasi *topic polarity wordcloud*. Adapun tujuan khusus penelitian ini adalah untuk mengetahui pengaruh perubahan nilai *alpha*, *beta*, jumlah topik terhadap nilai *perplexity* dan waktu eksekusi (detik), mengetahui kombinasi parameter terbaik untuk analisis sentimen, dan mengetahui persebaran topik terkait komentar mahasiswa Universitas Diponegoro. Pada penelitian ini dilakukan proses LDA untuk melakukan ekstraksi topik dengan menggunakan 690 data komentar. Kemudian memvisualisasikan hasil ekstraksi topik ke dalam bentuk *topic polarity wordcloud* dengan sentimen positif, negatif atau netral. Hasil penelitian menunjukkan nilai *alpha*, *beta*, dan jumlah topik yang semakin besar menghasilkan nilai *perplexity* yang kecil. Nilai *alpha* dan *beta* yang semakin besar menghasilkan waktu eksekusi yang relatif lebih cepat. Sedangkan kenaikan jumlah topik menghasilkan waktu eksekusi yang lebih lama. Kemudian kombinasi parameter terbaik adalah parameter dengan nilai *alpha* 0.1, *beta* 0.1, jumlah topik 9 dengan nilai *threshold* 0.0000001. Kombinasi parameter menghasilkan 3 topik sebagai sentimen positif dan 6 topik sebagai sentimen negatif dengan topik yang dibahas adalah masalah fasilitas, birokrasi, tempat parkir, dan ukt.

Kata kunci : *Latent Dirichlet Allocation*, analisis sentimen, *wordcloud*

ABSTRACT

For an educational institution, the opinions of learners have an important role for the progress of the educational institutions. Sentiment analysis can help in extracting the opinions from some documents, comments, product reviews, and other data that will be easier to do the monitoring. One method for doing sentiment analysis is Latent Dirichlet Allocation (LDA), which can be used to extract the topic of documents, where the topic is represented as the appearance of the words with different topic probability. Therefore, we need data representation in a visual form that is easier to understand than text and tables. One form of data visualization is wordcloud that provides a visual representation of words frequency. This research will perform sentiment analysis from a student's comments to an university, in this case the University of Diponegoro, using LDA and topic polarity wordcloud visualization. This study aims to determine the effect of changes of alpha, beta, the number of topics against the value of perplexity and execution time (seconds), determine the best combination of parameters for sentiment analysis, and determine the distribution of related topics from Diponegoro University student's comments. In this research performed LDA process for extracting topic using 690 dataset and then visualized to topic polarity wordcloud with sentiment positive, negative or neutral. The results show that the increase value of alpha, beta, and the number of topics produce smaller value of perplexity. Greater value of alpha and beta produce a faster execution time. But greater value of the number of topics produce slower execution time. The best parameter combination is parameter with value of alpha 0.1, value of beta 0.1, number of topics 9, threshold 0.0000001. This best parameter combination produces 3 topics as positive sentiment and 6 topics as negative sentiment with the topics are talking about facilities, bureaucratic, parking lots, and tuition.

Keywords : *Latent Dirichlet Allocation, sentiment analysis, wordcloud*

KATA PENGANTAR

Segala puji syukur bagi Allah SWT atas karunia-Nya yang diberikan kepada penulis sehingga penulis dapat menyelesaikan tugas akhir ini. Tugas akhir yang berjudul “Analisis Sentimen Menggunakan *Latent Dirichlet Allocation* dan Visualisasi *Topic Polarity Wordcloud*”. 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 tugas akhir ini tentulah telah banyak mendapat bantuan dan dukungan dari berbagai pihak. Untuk itu, pada kesempatan ini penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Bapak Ragil Saputra, S.Si, M.Cs, selaku Ketua Departemen Ilmu Komputer / Informatika FSM Universitas Diponegoro Semarang.
2. Helmie Arif Wibawa, S.Si, M.Cs, selaku Koordinator Tugas Akhir Departemen Ilmu Komputer/Informatika FSM Universitas Diponegoro Semarang
3. Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku dosen Pembimbing yang telah meluangkan waktu dan berkenan memberikan bimbingan, arahan, masukan, serta motivasi yang sangat berharga dan fokus akan tujuan bagi penulis.
4. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu kelancaran penelitian ini, semoga Tuhan yang memberikan balasan yang lebih baik.

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 pada umumnya dan penulis pada khususnya.

Semarang, 3 Februari 2017

Mohammad Fajar Ainul Bashri

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.....	2
1.3. Tujuan dan Manfaat	3
1.4. Ruang Lingkup.....	3
1.5. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1. <i>Sentiment Analysis</i>	5
2.2. <i>Wordcloud</i>	6
2.3. <i>Latent Dirichlet Allocation (LDA)</i>	7
2.4. <i>Preprocessing</i>	11
2.4.1. Tokenisasi.....	11
2.4.2. <i>Stopword</i>	12
2.4.3. <i>Stemming</i>	12
2.5. Evaluasi.....	20
2.6. Pengembangan Perangkat Lunak	20
BAB III METODOLOGI PENELITIAN	23
3.1. Pengumpulan Data Komentar	24
3.2. <i>Preprocessing</i>	24
3.2.1. Tokenisasi.....	25
3.2.2. Menghapus <i>Stopword</i>	26
3.2.3. Melakukan <i>Stemming</i> Kata.....	27

3.2.4.	Pembentukan <i>Bag of Words</i>	32
3.3.	LDA <i>Collapsed Gibbs Sampling</i>	33
3.4.	Membentuk Visualisasi <i>Topic Polarity Wordcloud</i>	35
3.5.	Analisa dan Desain Sistem.....	40
3.5.1.	Analisis Sistem	40
3.5.1.1.	Deskripsi Sistem.....	40
3.5.1.2.	Kebutuhan Fungsional Sistem.....	41
3.5.1.3.	Kebutuhan Non-Fungsional Sistem	41
3.5.2.	Perancangan Sistem.....	41
3.5.2.1.	Pemodelan Data.....	41
3.5.2.1.1.	<i>Entity Relationship Diagram</i> (ERD).....	42
3.5.2.1.2.	Deskripsi Entitas	42
3.5.2.1.3.	Deskripsi Relasi antar Entitas	43
3.5.2.2.	Pemodelan Fungsi	43
3.5.2.2.1.	<i>Data Context Diagram</i> (DCD).....	43
3.5.2.2.2.	<i>Data Flow Diagram Level 1</i>	44
3.5.2.2.3.	<i>Data Flow Diagram Level 2</i> Proses 2	45
3.5.2.3.	Perancangan Antarmuka	46
3.5.2.3.1.	Antarmuka Halaman Dasbor	46
3.5.2.3.2.	Antarmuka Halaman Lihat Komentar	46
3.5.2.3.3.	Antarmuka Halaman <i>Login</i>	47
3.5.2.3.4.	Antarmuka Halaman <i>Form</i>	47
BAB IV	HASIL DAN PEMBAHASAN	49
4.1.	Hasil Pengembangan Sistem.....	49
4.1.1.	Lingkungan Implementasi	49
4.1.2.	Implementasi Antarmuka	49
4.2.	Skenario Pengujian Sistem	52
4.2.1.	Pengujian Fungsional Sistem.....	52
4.2.2.	Pengujian Kinerja Sistem	52
4.2.2.1.	Data	53
4.2.2.2.	Eksperimen.....	53
4.2.2.2.1.	Skenario Eksperimen 1.....	55
4.2.2.2.2.	Skenario Eksperimen 2.....	55

4.2.2.2.3. Skenario Eksperimen 3.....	55
4.2.2.2.4. Skenario Eksperimen 4.....	56
4.3. Hasil dan Analisis Sistem	56
4.3.1. Pengujian Fungsional Sistem.....	56
4.3.2. Skenario Eksperimen 1	56
4.3.3. Skenario Eksperimen 2.....	58
4.3.4. Skenario Eksperimen 3.....	59
4.3.5. Skenario Eksperimen 4.....	62
BAB V PENUTUP.....	69
5.1. Kesimpulan	69
5.2. Saran	69
DAFTAR PUSTAKA.....	70
LAMPIRAN - LAMPIRAN	72
Lampiran 1. Tampilan formulir pengisian komentar <i>online</i>	73
Lampiran 2. Perhitungan LDA <i>Collapse Gibbs Sampling</i>	74
Lampiran 3. Daftar Rencana Pengujian dan Hasil pengujian.....	113
Lampiran 4. Detail jumlah komentar masing-masing fakultas.....	116
Lampiran 5. Dataset Komentar.....	118
Lampiran 6. Kombinasi Parameter.....	139
Lampiran 7. Hasil Nilai <i>Perplexity</i> dan Waktu Eksekusi.....	140

DAFTAR GAMBAR

Gambar 2.1. Contoh <i>wordcloud</i>	7
Gambar 2.2. Alur pembangkitan teks dengan <i>topic modeling</i> (Blei, 2012)	8
Gambar 2.3. Visualisasi LDA sebagai generatif (Kusumaningrum, 2014).....	8
Gambar 2.4. Contoh LDA sebagai generatif (Blei, 2012).....	8
Gambar 2.5. Visualisasi LDA sebagai inferensi (Kusumaningrum, 2014)	9
Gambar 2.6. Contoh LDA sebagai inferensi (Blei, 2012)	9
Gambar 2.7. Model <i>waterfall</i> (Sommerville, 2007)	21
Gambar 3.1. Garis besar penyelesaian masalah	23
Gambar 3.2. Diagram alur <i>preprocessing</i>	24
Gambar 3.3. Diagram alur proses tokenisasi	25
Gambar 3.4. Contoh proses tokenisasi	25
Gambar 3.5. Diagram alur proses penghapusan <i>stopword</i>	26
Gambar 3.6. Contoh penghapusan <i>stopword</i>	26
Gambar 3.7. Diagram alur proses <i>stemming</i>	27
Gambar 3.8. Diagram alur subproses <i>stemmer</i> Sastrawi	27
Gambar 3.9. Diagram alur subproses <i>stemPluralWord</i>	28
Gambar 3.10. Diagram alur subproses <i>stemSingularWord</i>	29
Gambar 3.11. Diagram alur subproses <i>loopPengembalianAkhiran</i>	31
Gambar 3.12. Contoh hasil <i>stemming</i>	31
Gambar 3.13. Diagram alur proses pembentukan <i>bag of words</i>	32
Gambar 3.14. Contoh <i>bag of words</i>	32
Gambar 3.15. Diagram alur proses LDA <i>Collapsed Gibbs Sampling</i>	33
Gambar 3.16. Diagram alur proses perangkingan	34
Gambar 3.17. Diagram alur proses pembentukan visualisasi <i>topic polarity wordcloud</i>	37
Gambar 3.18. Diagram alur subproses cetak <i>wordcloud</i>	38
Gambar 3.19. Contoh visualisasi <i>topic polarity wordcloud</i>	40
Gambar 3.20. Entity Relationship Diagram (ERD).....	42
Gambar 3.21. Data Context Diagram (DCD)	44
Gambar 3.22. <i>Data Flow Diagram (DFD) Level 1</i>	44
Gambar 3.23. <i>Data Flow Diagram Level 2</i>	45
Gambar 3.24. Antarmuka halaman dasbor	46

Gambar 3.25. Antarmuka halaman Lihat Komentar	47
Gambar 3.26. Antarmuka halaman Lihat Komentar (Melihat komentar)	47
Gambar 3.27. Antarmuka halaman <i>Login</i>	48
Gambar 3.28. Antarmuka halaman <i>form</i>	48
Gambar 4.1. Antarmuka halaman <i>login</i>	50
Gambar 4.2. Antarmuka halaman dasbor	50
Gambar 4.3. Antarmuka halaman lihat komentar.....	51
Gambar 4.4. Antarmuka halaman lihat komentar (detail komentar).....	51
Gambar 4.5. Antarmuka halaman pengisian komentar	52
Gambar 4.6. Grafik jumlah komentar untuk setiap fakultas.....	53
Gambar 4.7. Skenario eksperimen.....	54
Gambar 4.8. Grafik rata-rata nilai <i>perplexity</i> untuk setiap nilai <i>alpha</i>	57
Gambar 4.9. Grafik rata-rata waktu eksekusi untuk setiap nilai <i>alpha</i>	57
Gambar 4.10. Grafik rata-rata nilai <i>perplexity</i> untuk setiap nilai <i>beta</i>	58
Gambar 4.11. Grafik rata-rata waktu eksekusi untuk setiap nilai <i>beta</i>	59
Gambar 4.12. Grafik rata-rata nilai <i>perplexity</i> untuk jumlah topik 2, 4, dan 6	60
Gambar 4.13. Grafik rata-rata nilai <i>perplexity</i> untuk jumlah topik 3, 6, dan 9	60
Gambar 4.14. Grafik rata-rata waktu eksekusi untuk jumlah topik 2,4, dan 6.....	61
Gambar 4.15. Grafik rata-rata waktu eksekusi untuk jumlah topik 3, 6, dan 9.....	61
Gambar 4.16. Grafik nilai <i>perplexity</i> untuk masing-masing parameter kombinasi.....	63
Gambar 4.17. Visualisasi <i>topic polarity wordcloud</i> untuk topik 1.....	64
Gambar 4.18. Visualisasi <i>topic polarity wordcloud</i> untuk topik 2.....	64
Gambar 4.19. Visualisasi <i>topic polarity wordcloud</i> untuk topik 3.....	65
Gambar 4.20. Visualisasi <i>topic polarity wordcloud</i> untuk topik 4.....	65
Gambar 4.21. Visualisasi <i>topic polarity wordcloud</i> untuk topik 5.....	66
Gambar 4.22. Visualisasi <i>topic polarity wordcloud</i> untuk topik 6.....	66
Gambar 4.23. Visualisasi <i>topic polarity wordcloud</i> untuk topik 7.....	67
Gambar 4.24. Visualisasi <i>topic polarity wordcloud</i> untuk topik 8.....	67
Gambar 4.25. Visualisasi <i>topic polarity wordcloud</i> untuk topik 9.....	68

DAFTAR TABEL

Tabel 2.1. Penelitian Lain Terkait Analisis Sentimen	6
Tabel 2.2. Bentuk <i>bag of words</i>	11
Tabel 2.3. Contoh daftar <i>stopword</i> bahasa Indonesia.....	12
Tabel 2.4. Kombinasi awalan-akhiran yang dilarang	14
Tabel 2.5. Aturan pemenggalan awalan algoritma <i>stemming</i> Nazief Adriani.....	16
Tabel 2.6. Modifikasi dan penambahan aturan pemenggalan awalan oleh Algoritma <i>stemming Confix Stripping</i>	17
Tabel 2.7. Daftar aturan <i>rulePrecedence</i>	17
Tabel 2.8. Modifikasi aturan pemenggalan awalan oleh Algoritma <i>stemming Enhanced Confix Stripping</i>	18
Tabel 2.9. Modifikasi aturan pemenggalan awalan dan penambahan aturan pemenggalan sisipan oleh algoritma <i>stemming modified Enhanced Confix Stripping</i>	19
Tabel 2.10. Penambahan dan modifikasi aturan pemenggalan awalan <i>stemmer</i> Sastrawi..	20
Tabel 3.1. Hasil Proses LDA <i>Collapsed Gibbs Sampling</i>	34
Tabel 3.2. Hasil pengambilan <i>n</i> kata tertinggi setelah pengurutan.....	35
Tabel 3.3. Tabel <i>Sentiment Lexicon</i>	36
Tabel 3.4. Kebutuhan fungsional	41
Tabel 3.5. Kebutuhan non-fungsional	41
Tabel L.2.1. <i>Bag of words</i>	74
Tabel L.2.2. <i>Vocabulary</i>	74
Tabel L.2.3. <i>Bag of words</i> setelah dilakukan <i>random</i> topik.....	74
Tabel L.2.4. Hasil proses <i>Gibbs Sampling</i>	103
Tabel L.2.5. Hasil perhitungan nilai probabilitas kata untuk masing-masing topik.....	111
Tabel L.2.6. Hasil proses <i>ranking</i>	112
Tabel L.3.1. Daftar Rencana Pengujian.....	113
Tabel L.3.2. Daftar Hasil Pengujian	114
Tabel L.4.1. Jumlah komentar masing-masing fakultas	116
Tabel L.5.1. Tabel dataset komentar	118
Tabel L.6.1. Tabel kombinasi parameter penelitian	139
Tabel L.7.1. Hasil Nilai <i>Perplexity</i> dan Waktu Eksekusi untuk nilai <i>threshold</i> 0.001.....	140
Tabel L.7.2. Hasil Nilai <i>Perplexity</i> dan Waktu Eksekusi untuk nilai <i>threshold</i> 0.0001....	141

Tabel L.7.3. Hasil Nilai <i>Perplexity</i> dan Waktu Eksekusi untuk nilai <i>threshold</i> 0.00001..	142
Tabel L.7.4. Hasil Nilai <i>Perplexity</i> dan Waktu Eksekusi untuk nilai <i>threshold</i> 0.000001	143
Tabel L.7.5. Hasil Nilai <i>Perplexity</i> dan Waktu Eksekusi untuk nilai <i>threshold</i> 0.0000001	144

BAB I

PENDAHULUAN

Bab pendahuluan membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan Tugas Akhir mengenai analisis sentimen menggunakan *Latent Dirichlet Allocation* dan visualisasi *topic polarity wordcloud*.

1.1. Latar Belakang

Pendapat atau opini saat ini memiliki peran besar untuk setiap tindakan manusia. Saat seseorang ingin membuat sebuah keputusan, orang tersebut ingin tahu bagaimana pendapat orang lain tentang apa yang akan dikerjakannya. Misalnya ketika seseorang hendak membeli sebuah produk dari toko *online*, orang tersebut akan melihat *review* pengguna lain untuk memutuskan apakah akan membeli produk tersebut atau tidak. Bagi institusi yang bergerak di bidang barang maupun jasa, sangat diperlukan pendapat konsumen mengenai barang atau jasa yang dihasilkannya.

Sentiment analysis (analisis sentimen) atau yang juga disebut sebagai *opinion mining* adalah salah satu bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi seseorang terhadap suatu entitas tertentu seperti produk, jasa, organisasi, individu, masalah, peristiwa, topik dan atribut lainnya (Liu, 2012). Analisis sentimen dapat digunakan untuk melakukan ekstraksi opini-opini dari dokumen, komentar, sosial media, *review blog* dan data-data lainnya. Analisis sentimen mempunyai beberapa manfaat seperti pemantauan terhadap sebuah produk hingga pemantauan terhadap tingkat kepuasan terhadap sistem pendidikan yang diterapkan.

Ada beberapa metode yang dapat digunakan untuk melakukan analisis sentimen, salah satu metode yang dapat digunakan adalah dengan menggunakan *probabilistic topic modeling*. *Probabilistic topic modeling* adalah serangkaian algoritma yang bertujuan untuk menemukan topik utama pada kumpulan dokumen berukuran besar dan tidak terstruktur (Blei, 2012). *Probabilistic topic modeling* memberikan kemudahan untuk mengorganisir dan meringkas arsip elektronik dalam skala yang besar. Saat ini metode *probabilistic topic modeling* yang terkenal adalah *Latent Dirichlet Allocation* (LDA) yang memiliki asumsi komprehensif pada

pembangkitan teks daripada metode yang lainnya, yaitu TF-IDF, Mixture Unigrams, LSA, pLSA. (Liu, 2013). Hal ini dikarenakan LDA memodelkan ϕ dan θ berdistribusi *Dirichlet* dari α dan β yang merupakan *conjugate prior* dari probabilitas kata. Sehingga mengakibatkan parameter tidak bertumbuh linier terhadap jumlah data.

Selanjutnya, tahapan merepresentasikan data dalam bentuk visual lebih memudahkan dalam memahami isi data daripada data dalam bentuk teks atau tabel. Visualisasi membantu orang dalam memahami data yang susah dipahami. Bahkan ketika jumlah data yang digunakan sangat besar, pola data dapat dipahami dengan cepat dan mudah. Visualisasi data menyampaikan informasi dengan cara yang universal dan membuatnya lebih sederhana. Salah satu bentuk visualisasi data adalah *wordcloud*. *Wordcloud* adalah sebuah gambaran visual dari frekuensi kemunculan kata-kata pada buku teks atau situs *internet*. Ukuran huruf menentukan frekuensi kemunculan sebuah kata, maka semakin besar ukuran huruf maka semakin besar kemunculan kata tersebut. Sebaliknya, semakin kecil ukuran huruf maka semakin kecil frekuensi kemunculan kata tersebut (Miley & Read, 2011).

Saat ini *wordcloud* masih dihasilkan berdasarkan frekuensi kemunculan kata pada dokumen dan tidak bisa mengakomodasi hubungan kemunculan kata dengan topik. Selain itu *wordcloud* belum bisa menunjukkan *polarity* dari isi dokumen. Oleh karena itu diperlukan sebuah metode lain yang dapat menghasilkan *wordcloud* dengan memunculkan kata-kata yang merepresentasikan topik dan *polarity* pada dokumen. Pada penelitian ini akan dilakukan visualisasi analisis sentimen dengan menggunakan visualisasi *polarity wordcloud*. Proses implementasi analisis sentimen pada penelitian menggunakan data komentar mahasiswa Universitas Diponegoro dalam menilai Universitas Diponegoro sebagai sebuah organisasi. Adapun data komentar tersebut merupakan data primer yang diperoleh dengan menyebarkan kuesioner *online*.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan dapat disusun rumusan masalah yaitu bagaimana menerapkan metode *Latent Dirichlet Allocation* pada bidang analisis sentimen dengan menggunakan visualisasi *topic polarity wordcloud*

untuk menilai opini mahasiswa terhadap suatu universitas, khususnya Universitas Diponegoro.

1.3. Tujuan dan Manfaat

Tujuan umum yang ingin dicapai dari penelitian Tugas Akhir ini adalah menghasilkan visualisasi *topic polarity wordcloud* dari analisis sentimen menggunakan *Latent Dirichlet Allocation* (LDA) untuk menilai opini mahasiswa terhadap universitas. Secara lebih mendetail, tujuan tersebut dapat dijabarkan ke dalam tujuan khusus, yaitu :

1. Mengetahui pengaruh perubahan nilai alpha dan beta terhadap nilai *perplexity* dan waktu eksekusi.
2. Mengetahui pengaruh perubahan jumlah topik terhadap nilai *perplexity* dan waktu eksekusi.
3. Mengetahui kombinasi parameter terbaik untuk analisis sentimen komentar mahasiswa Universitas Diponegoro.
4. Mengetahui persebaran topik terkait opini mahasiswa Universitas Diponegoro.

Adapun manfaat yang diharapkan dari penelitian Tugas Akhir ini adalah dapat digunakan sebagai bahan evaluasi institusi yang dalam hal ini adalah Universitas Diponegoro.

1.4. Ruang Lingkup

Ruang lingkup dalam menerapkan metode *Latent Dirichlet Allocation* dalam menggunakan visualisasi *topic polarity wordcloud* adalah sebagai berikut :

1. Instansi yang digunakan sebagai objek penelitian adalah Universitas Diponegoro.
2. Komentar diperoleh dari mahasiswa Universitas Diponegoro program Diploma III dan Strata I.
3. Jumlah komentar adalah 10 komentar untuk masing-masing program studi.
4. Komentar menggunakan bahasa Indonesia baku tanpa menggunakan singkatan, maksimal berjumlah 50 kata dan tidak menggunakan *emoticon*.

1.5. Sistematika Penulisan

Sistematika yang digunakan dalam laporan Tugas Akhir ini terbagi menjadi beberapa pokok bahasan, yaitu:

BAB I PENDAHULUAN

Bab pendahuluan membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan Tugas Akhir mengenai analisis sentimen menggunakan *Latent Dirichlet Allocation* dan visualisasi *topic polarity wordcloud*.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan tentang keseluruhan dari teori-teori yang digunakan dalam pengerjaan penelitian analisis sentimen menggunakan *Latent Dirichlet Allocation* dan visualisasi *topic polarity wordcloud* yaitu, *sentiment analysis*, *wordcloud*, *Latent Dirichlet Allocation*, *preprocessing*, evaluasi, dan pengembangan perangkat lunak.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan garis besar penyelesaian masalah tugas akhir yang diawali dengan penyajian garis besar penyelesaian masalah dalam bentuk blok proses.

BAB IV HASIL EKSPERIMEN DAN ANALISIS

Bab ini menjelaskan hasil eksperimen yang telah dilakukan. Pada bab ini akan dijelaskan mengenai data penelitian, skenario eksperimen, hasil eksperimen dan hasil visualisasi *topic polarity wordcloud*.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari pengerjaan penelitian ini, beserta dengan saran yang dapat diajukan guna pengembangan penelitian ke depannya.