

***SENTIMENT ANALYSIS PADA MEDIA SOSIAL FOURSQUARE
MENGUNAKAN SENTIMENT LEXICON DAN METODE
LATENT DIRICHLET ALLOCATION (LDA)***



SKRIPSI

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

Disusun oleh :

DARMAWAN NUR KUSUMA

24010313120039

**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO**

2017

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Darmawan Nur Kusuma

NIM : 24010313120039

Judul : *Sentiment Analysis* pada Media Sosial *Foursquare* Menggunakan *Sentiment Lexicon* dan Metode *Latent Dirichlet Allocation (LDA)*

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.

Semarang, 29 Desember 2017



Darmawan Nur Kusuma
NIM. 24010313120039


HALAMAN PENGESAHAN

Judul : *Sentiment Analysis* pada Media Sosial *Foursquare* Menggunakan
Sentiment Lexicon dan Metode *Latent Dirichlet Allocation* (LDA)
Nama : Darmawan Nur Kusuma
NIM : 24010313120039

Telah diujikan pada sidang tugas akhir pada tanggal 15 Desember 2017 dan dinyatakan lulus pada tanggal 15 Desember 2017.

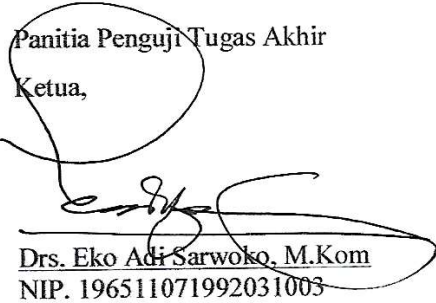
Semarang, 29 Desember 2017

Mengesahkan
Ketua Departemen Ilmu Komputer/ Informatika



Darmawan Kusumaningrum, S.Si. M.Kom
NIP. 198104202005012001

Panitia Penguji Tugas Akhir
Ketua,



Drs. Eko Adi Sarwoko, M.Kom
NIP. 196511071992031003

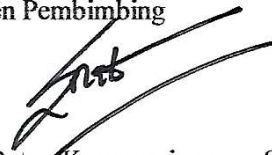
HALAMAN PENGESAHAN

Judul : *Sentiment Analysis* pada Media Sosial *Foursquare* Menggunakan
Sentiment Lexicon dan Metode *Latent Dirichlet Allocation* (LDA)
Nama : Darmawan Nur Kusuma
NIM : 24010313120039

Telah diujikan pada sidang tugas akhir pada tanggal 15 Desember 2017.

Semarang, 29 Desember 2017

Dosen Pembimbing



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

ABSTRAK

Foursquare merupakan salah satu media sosial yang dapat digunakan sebagai sarana untuk mencari letak restoran melalui *Global Positioning System* (GPS) sekaligus penggunaannya dapat memberikan opini berupa komentar pada restoran tersebut. Opini yang dituliskan berupa penilaian atau ulasan yang dikemas dalam bentuk komentar yang bermakna positif ataupun negatif. Opini dapat membantu pengguna atau pembeli untuk dapat memilih restoran yang *recommended* dan juga dapat membantu penjual untuk mengetahui perkembangan restoran miliknya. Opini dalam jumlah yang terlalu banyak, terkadang membuat penjual kesulitan untuk menganalisisnya. Penelitian ini bertujuan untuk menjawab permasalahan tersebut dengan menganalisa opini yang ada melalui *sentiment analysis*. *Sentiment analysis* dengan *probabilistic topic model* menggunakan metode *Latent Dirichlet Allocation* (LDA) merupakan teknologi yang dapat diimplementasikan untuk membaca opini publik terkait sentimen positif dan negatif terhadap topik tertentu. Penelitian yang diusulkan merupakan *sentiment analysis* pada *Foursquare* menggunakan *sentiment lexicon* dan metode *Latent Dirichlet Allocation* (LDA). Hasil dari beberapa eksperimen pada penelitian ini menunjukkan bahwa penggunaan normalisasi *Jaro Winkler Distance* dapat meningkatkan akurasi sebesar 3,34 %. Ekstraksi fitur menggunakan *sentiment term frequency* menghasilkan akurasi lebih tinggi daripada *sentiment term frequency* dengan selisih akurasi sebesar 1,34 %. Eksperimen dengan melakukan perubahan jumlah topik mendapatkan nilai akurasi tertinggi pada ekstraksi fitur *sentiment term occurrence*, yang menghasilkan nilai akurasi sebesar 79,50 % dengan jumlah topik optimal sebesar 300.

Kata kunci: *Sentiment Analysis, Foursquare, Restoran, Latent Dirichlet Allocation*

ABSTRACT

Foursquare is one of the social media which be used as a tool to find the location of a restaurant by using Global Positioning System (GPS), it also provide any comment about opinion of that restaurant. The opinion written as an assessment or review of it, written either positive or negative comment. The opinion help users or buyers to choose a recommended restaurant and also help the owners of restaurant to identify the progress of theirs. In the other hand the owner sometimes find difficulties to analyze the opinions. This research aims to provide solution by analyzing the opinions through sentiment analysis. Sentiment analysis with probabilistic topic model using Latent Dirichlet Allocation (LDA) method is a technology can be implemented to read public opinion related to positive and negative sentiments of a certain topic. The goal of this research is sentiment analysis on Foursquare using lexicon sentiment and Latent Dirichlet Allocation (LDA) method. The results of experiments indicate that application of Jaro Winkler Distance normalization can increase accuracy by 3.34%. Feature extraction by using sentiment term occurrence yields higher accuracy than sentiment term frequency with accuracy difference is 1.34%. The experiment with changing the number of topics results the highest accuracy value on feature extraction sentiment term occurrence, the result is of 79.50% with an optimum topic count of is 300.

Keywords: Sentiment Analysis, Foursquare, Restaurant, Latent Dirichlet Allocation

KATA PENGANTAR

Puji syukur kepada Allah SWT yang telah memberikan rahmat dan hidayahnya-Nya sehingga penulisan dapat menyelesaikan skripsi dengan judul “*Sentiment Analysis* pada Media Sosial *Foursquare* Menggunakan *Sentiment Lexicon* dan Metode *Latent Dirichlet Allocation* (LDA)” dengan baik dan lancar.

Tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu (S1) pada Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Pada penyusunan tugas akhir ini, penulis banyak mendapat bimbingan, arahan dan bantuan dari berbagai pihak. Oleh karena itu, melalui kesempatan kali ini penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Ibu Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku Ketua Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang, sekaligus menjadi Dosen Pembimbing yang telah membantu dalam proses bimbingan hingga selesainya laporan tugas akhir ini.
2. Bapak Helmie Arif Wibawa, S.Si, M.Cs, selaku Koordinator Tugas Akhir Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.
3. Bapak Sarwoto dan Ibu Darini, selaku orang tua yang terus memberikan dukungan dan do’a dalam menyelesaikan tugas akhir ini.
4. Semua pihak yang telah membantu kelancaran dalam penyusunan tugas akhir, yang tidak dapat disebutkan satu persatu.

Pada pembuatan laporan tugas akhir ini masih memiliki banyak kekurangan baik dari segi materi ataupun dari segi penyajiannya karena keterbatasan kemampuan dan pengetahuan dari penulis. Oleh karena itu, kritik dan saran sangat diharapkan. Terima kasih, dan semoga laporan tugas akhir ini bermanfaat bagi semua pihak.

Semarang, 4 Desember 2017

Darmawan Nur Kusuma

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI	ii
HALAMAN PENGESAHAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
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>Sentiment Lexicon</i>	6
2.3. <i>Latent Dirichlet Allocation (LDA)</i>	6
2.3.1. Proses Pelatihan <i>Latent Dirichlet Allocation (LDA)</i>	8
2.3.2. Proses Pengujian	10
2.4. <i>K-Fold Cross Validation</i>	10
2.5. <i>Preprocessing</i>	11
2.5.1. <i>Case folding</i>	11
2.5.2. Tokenisasi	11
2.5.3. Normalisasi	12
2.5.4. <i>Filtering</i>	14
2.5.5. <i>Stemming</i>	14
2.6. <i>Kullback Leibler Divergence</i>	23
2.7. Evaluasi	24

BAB III METODOLOGI PENELITIAN.....	26
3.1. Pengumpulan Data.....	27
3.2. <i>Preprocessing</i>	28
3.2.1. <i>Case Folding</i>	29
3.2.2. Tokenisasi.....	29
3.2.3. Normalisasi.....	31
3.2.4. <i>Filtering</i>	34
3.2.5. <i>Stemming</i>	36
3.3. Ekstraksi Fitur.....	45
3.4. <i>K-Fold Cross Validation</i>	46
3.5. Pembentukan Model <i>Latent Dirichlet Allocation</i> (LDA).....	48
3.5.1. Pelatihan.....	48
3.5.2. Pengujian.....	54
3.6. Validasi dan Perhitungan Nilai Akurasi.....	56
3.7. Pembentukan Aplikasi <i>Sentiment Anlalysis</i> Restoran.....	58
3.8. Analisis dan Desain Aplikasi.....	59
3.8.1. Analisis Aplikasi.....	59
3.8.2. Perancangan Aplikasi.....	60
BAB IV HASIL DAN ANALISA.....	68
4.1. Hasil Pengembangan Apikasi.....	68
4.1.1. Lingkungan Implementasi.....	68
4.1.2. Implementasi Proses.....	68
4.1.3. Implementasi Antarmuka.....	69
4.2. Skenario Pengujian Aplikasi.....	74
4.2.1. Skenario Pengujian Fungsional Aplikasi.....	74
4.2.2. Skenario Pengujian Kinerja Aplikasi.....	75
4.3. Hasil dan Analisa Aplikasi.....	79
4.3.1. Hasil dan Analisa Pengujian Fungsional Aplikasi.....	79
4.3.2. Hasil dan Analisa Pengujian Kinerja Aplikasi.....	79
BAB V PENUTUP.....	86
5.1. Kesimpulan.....	86
5.2. Saran.....	87
DAFTAR PUSTAKA.....	88

DAFTAR GAMBAR

Gambar 2.1 Graphical Method of LDA (Kusumaningrum, et al., 2014).....	6
Gambar 2.2 Contoh LDA sebagai Generatif (Blei, 2012)	7
Gambar 2.3 Visualisasi LDA sebagai Inferensi (Kusumaningrum, et al., 2014).....	8
Gambar 2.4 Tahap Preprocessing	11
Gambar 3.1 Gambaran Umum Penelitian	26
Gambar 3.2 Flowchart Preprocessing	28
Gambar 3.3 Flowchart Tokenisasi	31
Gambar 3.4 Flowchart Normalisasi	31
Gambar 3.5 Flowchart Stopword Removal.....	34
Gambar 3.6 Proses Stemming	36
Gambar 3.7 Stemming PluralWord (Bashri, 2017)	37
Gambar 3.8 Sub-proses stemming sastrawi (Bashri, 2017)	37
Gambar 3.9 Stemming SingularWord (Bashri, 2017)	38
Gambar 3.10 loopPengembalianAkhiran (Bashri, 2017).....	40
Gambar 3.11 Pembagian dan Pengindeksan data (K-Fold CV).....	47
Gambar 3.12 Flowchart Pembentukan Model LDA.....	48
Gambar 3.13 Alur proses pelatihan	49
Gambar 3.14 Proses LDA Collapsed Gibbs Sampling.....	50
Gambar 3.15 Flowchart Pengujian	56
Gambar 3.16 Data Context Diagram (DCD).....	61
Gambar 3.17 Data Flow Diagram level 1 (DFD).....	61
Gambar 3.18 Data Flow Diagram level 2 Preprocessing.....	62
Gambar 3.19 Antarmuka Halaman Masukan Data Restoran	63
Gambar 3.20 Antarmuka Halaman Hasil Analisa	64
Gambar 3.21 Antarmuka Halaman Data Pemodelan.....	65
Gambar 3.22 Antarmuka Halaman Sentiment Lexicon.....	65
Gambar 3.23 Antarmuka Halaman Preprocessing.....	66
Gambar 3.24 Antarmuka Halaman Masukan Parameter	66
Gambar 3.25 Antarmuka Halaman Hasil Pemodelan	67
Gambar 4.1 Implementasi Halaman Masukan Data Restoran	70
Gambar 4.2 Implementasi Halaman Hasil Analisa.....	70

Gambar 4.3 Implementasi Halaman Data Pemodelan	71
Gambar 4.4 Implementasi Halaman Sentiment Lexicon	71
Gambar 4.5 Implementasi Halaman Preprocessing.....	72
Gambar 4.6 Implementasi Halaman Masukan Parameter.....	73
Gambar 4.7 Implementasi Halaman Hasil Pemodelan	73
Gambar 4.8 Rincian Pembagian Data Pemodelan.....	75
Gambar 4.9 Perbandingan hasil akurasi pada preprocessing dengan normalisasi dan tanpa normalisasi	79
Gambar 4.10 Grafik Hasil Eksperimen (Sentiment Term Occurrence).....	81
Gambar 4.11 Nilai Fold dari Kombinasi S2-K16.....	81
Gambar 4.13 Nilai Fold dari Kombinasi S3-K14.....	83
Gambar 4.12 Grafik Hasil Eksperimen (Sentiment Term Frequency)	83
Gambar 4.15 Perbandingan Sentiment Term Occurrence dan Sentiment Term Frequency	84
Gambar 4.16 Grafik Perbandingan Sentiment Term Occurrence dan Sentiment Term Frequency dengan Jumlah Topik Kelipatan 20.....	85

DAFTAR TABEL

Tabel 2.1 Penelitian Lain Terkait <i>Sentiment Analysis</i>	5
Tabel 2.2 Keterangan Notasi dan Definisi Persamaan 2.1 sampai Persamaan 2.3.....	9
Tabel 2.3 Contoh Proses Tokenisasi.....	12
Tabel 2.4 Contoh Proses Normalisasi.....	13
Tabel 2.5 Contoh Proses Filtering	14
Tabel 2.6 Kombinasi awalan-akhiran yang dilarang	16
Tabel 2.7 Aturan pemenggalan awalan algoritma <i>stemming</i> Nazief Adriani	17
Tabel 2.8 Modifikasi dan penambahan aturan pemenggalan awalan oleh Algoritma stemming Confix Stripping	19
Tabel 2.9 Daftar aturan RulePrecedence.....	19
Tabel 2.10 Modifikasi aturan pemenggalan awalan oleh Algoritma stemming Enhanced Confix Stripping.	20
Tabel 2.11 Modifikasi aturan pemenggalan awalan dan penambahan aturan pemenggalan sisipan oleh algoritma stemming modified Enhanced Confix Stripping	21
Tabel 2.12 Penambahan dan modifikasi aturan pemenggalan awalan <i>stemmer</i> Sastrawi ...	21
Tabel 2.13 Contoh kata dan nilai m untuk Porter Stemmer.....	22
Tabel 2.14 Ilustrasi Hasil Proses Stemming Menggunakan Porter Stemmer	23
Tabel 2.15 Confusion matrix untuk mencari nilai accuracy	24
Tabel 3.1 Dataset Komentar Sebelum Case Folding	29
Tabel 3.2 Dataset Komentar Setelah Case Folding	29
Tabel 3.3 Dataset Komentar Sebelum Tokenisasi.....	30
Tabel 3.4 Dataset Komentar Setelah Tokenisasi.....	30
Tabel 3.5 Dataset Komentar Sebelum Normalisasi.....	32
Tabel 3.6 Dataset Komentar Setelah Normalisasi	32
Tabel 3.7 Contoh beberapa stopwords Bahasa Inggris.....	35
Tabel 3.8 Contoh beberapa stopwords Bahasa Indonesia.....	35
Tabel 3.9 Contoh Stemming SingularWord.....	39
Table 3.10 Alur Proses Stemming Porter.....	41
Tabel 3.11 Rule Condition Pada Step 1	42
Tabel 3.12 Rule Condition Pada Step 2a	42
Tabel 3.13 Rule Condition Pada Step 2b	42

Tabel 3.14 Rule Condition Pada Step 3	43
Tabel 3.15 Rule Condition Pada Step 4	43
Table 3.16 Rule Condition Pada Step 5	44
Table 3.17 Rule Condition Pada Step 6	44
Table 3.18 Rule Condition Pada Step 7a	45
Table 3.19 Rule Condition Pada Step 7b	45
Tabel 3.20 Contoh Sentiment Lexicon	45
Tabel 3.21 Perbandingan Ekstraksi Fitur	46
Tabel 3.22 Representasi BoW	47
Tabel 3.23 Kombinasi Parameter Eksperimen	51
Tabel 3.24 Contoh Vocabulary	52
Tabel 3.25 Nilai topic proportion (PZD).....	53
Tabel 3.26 Hasil Perhitungan Nilai PZC	54
Tabel 3.27 Contoh topic proportion (PZD) uji.....	55
Tabel 3.28 Contoh topic proportion per kelas (PZC) latih.....	55
Tabel 3.29 Contoh Dataset Asli dan Dataset prediksi	57
Tabel 3.30 Contoh perhitungan confusion matrix	58
Tabel 3.31 Kebutuhan Fungsional	60
Tabel 3.32 Kebutuhan Non-Fungsional	60
Tabel 4.1 Skenario Pengujian Fungsional Aplikasi.....	74
Tabel 4.2 Hasil Skenario 2 (<i>Sentiment Term Occurrence</i>)	80
Tabel 4.3 Hasil Skenario 3 (<i>Sentiment Term Frequency</i>)	82
Tabel 6.1 Contoh Bag of Word (BoW).....	92
Tabel 6.2 Contoh Vocabulary.....	92
Tabel 6.3 Inisialisasi Topik Random	93
Tabel 6.4 Iterasi 1	93
Tabel 6.5 Iterasi 1 Token 1.....	94
Tabel 6.6 Iterasi 1 Token 2.....	96
Tabel 6.7 Iterasi 1 Token 3.....	97
Tabel 6.8 Iterasi 1 Token 4.....	98
Tabel 6.9 Iterasi 1 Token 5.....	99
Tabel 6.10 Iterasi 1 Token 6.....	100
Tabel 6.11 Iterasi 1 Token 7.....	102

Tabel 6.12 Iterasi 1 Token 8.....	103
Tabel 6.13 Iterasi 1 Token 9.....	104
Tabel 6.14 Probabilitas Kata Topik (PWZ)	105
Tabel 6.15 Topic Proportion (PZD).....	105
Tabel 6.16 Probabilitas Topik	106

DAFTAR LAMPIRAN

Lampiran 1.	Perhitungan Metode <i>Latent Dirichlet Allocation (LDA) Collapsed Gibbs Sampling</i>	92
Lampiran 2.	Source Code Implementasi Proses.....	107
Lampiran 3.	<i>Dataset</i> Komentar Restoran di Semarang dari <i>Foursquare</i>	128
Lampiran 4.	Hasil Pengujian Fungsional	139
Lampiran 5.	Hasil Akurasi Menggunakan Normalisasi <i>Jaro Winkler Distance</i>	141
Lampiran 6.	Hasil Akurasi Tanpa Menggunakan Normalisasi <i>Jaro Winkler Distance</i>	142
Lampiran 7.	Hasil Akurasi Eksperimen Jumlah Topik Kelipatan 20 Menggunakan Ekstraksi Fitur <i>Sentiment Term Occurrence</i>	143
Lampiran 8.	Hasil Akurasi Eksperimen Jumlah Topik Kelipatan 20 Menggunakan Ekstraksi Fitur <i>Sentiment Term Frequency</i>	147

BAB I

PENDAHULUAN

Bab pendahuluan membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat, serta ruang lingkup pelaksanaan tugas akhir *sentiment analysis* pada media sosial *Foursquare* menggunakan *sentiment lexicon* dan metode *Latent Dirichlet Allocation (LDA)*.

1.1. Latar Belakang

Seiring dengan perkembangan waktu, mulai bermunculan media sosial yang sering digunakan oleh masyarakat. Media sosial merupakan tempat dimana seseorang dapat melakukan *sharing* tentang aktivitas atau kegiatan yang sedang dilakukan, salah satunya adalah *Foursquare*. Sebagai media sosial yang berpusat pada lokasi dengan memanfaatkan *Global Positioning System (GPS)*, *Foursquare* menyimpan informasi tentang berbagai tempat dan telah mencatat sebagian besar opini pengguna untuk lokasi tersebut. Hal tersebut menyebabkan *Foursquare* dapat menarik puluhan juta pengguna di seluruh dunia (Chen, et al., 2016). Salah satu kategori yang paling diminati adalah restoran. Opini yang dituliskan pada *Foursquare* berupa penilaian atau ulasan yang dikemas dalam bentuk komentar. Komentar tersebut dapat membantu pengguna dalam memilih restoran yang *recommended* dan membantu penjual untuk melihat bagaimana tanggapan pengunjung terhadap restoran miliknya. Tentunya dengan tanggapan yang positif, dapat membuka peluang untuk meningkatkan keuntungan suatu restoran. Begitu juga sebaliknya, jika didapatkan tanggapan negatif maka perlu dilakukan perbaikan. Komentar dalam jumlah yang terlalu banyak, terkadang membuat penjual kesulitan untuk menganalisisnya. Maka dari itu diperlukan pengolahan data dan analisa terhadap komentar yang ada dengan menerapkan *sentiment analysis*.

Sentiment analysis, atau disebut juga dengan *opinion mining* merupakan bidang studi yang menganalisa pendapat setiap orang, sentimen, evaluasi, penilaian, sikap, dan perasaan terhadap entitas seperti produk, jasa organisasi, individu, masalah, peristiwa, topik, dan atribut-atributnya (Liu, 2012). *Sentiment analysis* dapat digunakan untuk mengkategorikan pendapat berdasarkan sentimennya. *Sentiment analysis* pada penelitian ini menggunakan restoran sebagai objek penelitian, sehingga dapat dicapai dengan memperhatikan cita rasa makanan, suasana, ragam makanan, harga, dan pelayanan. Sentimen yang diperhatikan adalah sentimen positif dan sentimen negatif.

Terdapat beberapa permasalahan yang harus diselesaikan pada saat melakukan *sentiment analysis*. Permasalahan yang muncul adalah sering ditemukannya kata-kata yang tidak sesuai dengan bahasa baku pada data-data opini atau terdapat penekanan pada sentimen yang berlebihan sehingga sulit untuk dianalisa. Hal tersebut mengakibatkan perlu adanya proses normalisasi untuk memperbaiki kesalahan penulisan kata yang tidak sesuai dengan kata baku. Normalisasi dapat dilakukan dengan menggunakan algoritma *Jaro Winkler Distance*, dimana algoritma ini bekerja dengan cara mengukur kesamaan antara dua *string*. Algoritma ini sering digunakan untuk pendeteksian duplikat (Rochmawati & Kusumaningrum, 2016).

Hal lain yang perlu diperhatikan pada *sentiment analysis* adalah pemilihan fitur yang efektif untuk digunakan pada proses *sentiment classification*. Beberapa contoh fiturnya adalah kata dan frekuensi katanya, *part of speech*, kata sentimenn, *rules of opinions*, *sentiment shifter* dan *syntactic dependency* (Liu, 2012). Pada penelitian ini menggunakan fitur kata sentiment yang dibedakan menjadi 2 jenis, yaitu frekuensi kata sentimen (*sentiment term frequency*) dan kemunculan kata sentimen (*sentiment term occurrence*). Pemilihan fitur tersebut dilakukan karena dapat memberikan hasil akurasi yang lebih baik (Azhari & Hidayatullah, 2014). Kedua ekstraksi fitur tersebut membutuhkan suatu kamus kata yang berisikan kata-kata sentimen, maka perlu adanya *sentiment lexicon* sebagai kamus kumpulan kata-kata sifat bernilai sentimen positif dan kata-kata sifat bernilai sentimen negatif.

Penelitian yang berkaitan dengan *sentiment analysis* suatu restoran sudah pernah dilakukan oleh Utami dan Wahono (2015) dengan menggunakan metode *Naïve Bayes*. Pada penelitian tersebut mengatakan bahwa menggunakan metode *Naïve Bayes* masih memerlukan beberapa penambahan untuk meningkatkan akurasi berdasarkan jumlah fitur yang ada. Selain itu *sentiment analysis* tentang *Foursquare* sudah pernah dilakukan oleh Aliandu (2015) menggunakan metode *Naïve Bayes*. Pada penelitian tersebut mengatakan bahwa *Naïve Bayes* memiliki kelemahan yaitu memiliki independensi fitur. Penggunaan metode lain digunakan oleh Prasidya (2016) pada penelitian *sentiment analysis* menggunakan *Bayesian Networks*. Salah satu dari masalah yang ditimbulkan oleh *Bayesian Networks* yaitu kesulitan komputasi untuk mengeksplorasi jaringan yang sebelumnya tidak dikenal. Perhitungan probabilitas dari setiap cabang jaringan harus dihitung, hal tersebut menyebabkan *cost* yang besar. Pada penelitian ini diberikan usulan menggunakan metode LDA yang merupakan salah satu model *Bayesian*. Metode LDA merupakan model *generative probabilistic* untuk koleksi data diskrit seperti korpus (Blei, 2003). LDA

memiliki keunggulan yaitu cocok untuk data dalam jumlah besar dan lebih stabil karena LDA menggunakan model hierarki (Liu, 2013).

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka dapat dirumuskan permasalahan yaitu bagaimana menerapkan *sentiment lexicon* dan metode LDA untuk melakukan *sentiment analysis* terhadap komentar pada media sosial *Foursquare*.

1.3. Tujuan dan Manfaat

Tujuan secara umum dari penelitian tugas akhir ini adalah melakukan *sentiment analysis* terhadap komentar pada media sosial *Foursquare* menggunakan *sentiment lexicon* dan metode LDA. Tujuan khusus dari penelitian tugas akhir ini, antara lain:

1. Mengetahui kinerja penerapan normalisasi pada *preprocessing* dengan menggunakan *Jaro Winkler Distance*.
2. Mengetahui kinerja ekstraksi fitur dengan menggunakan *sentiment term occurrence*.
3. Mengetahui kinerja ekstraksi fitur dengan menggunakan *sentiment term frequency*.
4. Membandingkan hasil dari kinerja ekstraksi fitur *sentimen term occurrence* dan *sentimen term frequency*.
5. Mengetahui jumlah topik yang paling optimal untuk menghasilkan nilai akurasi tertinggi.

Adapun manfaat yang diharapkan dari penelitian tugas akhir ini yaitu menghasilkan *sentiment analysis* yang dapat dimanfaatkan lebih lanjut untuk membantu dalam mengetahui sentimen publik terhadap suatu restoran bagi pengunjung ataupun pengelola restoran melalui komentar di media sosial *Foursquare*.

1.4. Ruang Lingkup

Ruang lingkup dalam menerapkan *sentiment analysis* menggunakan *sentiment lexicon* dan metode LDA adalah sebagai berikut:

1. Objek penelitian yang digunakan pada tugas akhir ini adalah restoran di kota Semarang.
2. Data diperoleh dari *crawling* komentar pada media sosial *Foursquare* sebanyak 300 data komentar berupa data dengan jumlah seimbang untuk beberapa restoran

yang terdiri dari 150 data sentimen positif dan 150 data sentimen negatif.

3. Menggunakan nama restoran dan id restoran yang ada di media sosial *Foursquare* untuk melakukan *sentiment analysis*.

1.5. Sistematika Penulisan

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

BAB I PENDAHULUAN

Bab pendahuluan membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat, serta ruang lingkup pelaksanaan tugas akhir *sentiment analysis* pada media sosial *Foursquare* menggunakan *sentiment lexicon* dan metode *Latent Dirichlet Allocation (LDA)*.

BAB II TINJAUAN PUSTAKA

Bab ini membahas mengenai kajian pustaka yang berhubungan dengan tugas akhir sebagai landasan untuk merumuskan dan menganalisa permasalahan pada tugas akhir. Kajian pustaka yang digunakan meliputi *sentiment analysis*, *sentiment lexicon*, *LDA*, *preprocessing*, *Kullback Leibler Divergence* dan evaluasi.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan yang dilakukan dalam penelitian tugas akhir. Tahapan tersebut meliputi pengumpulan data, *preprocessing*, ekstraksi fitur, *K-Fold*, pembentukan model, evaluasi dan proses *sentiment analysis* restoran.

BAB IV HASIL DAN ANALISA

Bab ini menguraikan hasil skenario dan analisa eksperimen yang dimulai dari teknis pengumpulan data sampai hasil dan analisa dari setiap eksperimen yang dilakukan.

BAB V PENUTUP

Bab ini menjelaskan mengenai kesimpulan dari uraian yang telah dijabarkan pada bab-bab sebelumnya dan saran untuk pengembangan penelitian lebih lanjut.