

HIERARCHICAL SENTENCE SENTIMENT ANALYSIS
UNTUK ULASAN HOTEL PADA *WEBSITE* TRAVELOKA
MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*



SKRIPSI

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

Disusun oleh:
Sandy Kurniawan
24010314120009

DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
2018

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Sandy Kurniawan

NIM : 24010314120009

Judul : *Hierarchical Sentence Sentiment Analysis* untuk Ulasan Hotel pada *Website* Traveloka Menggunakan Metode *Naïve Bayes Classifier*

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang sepengetahuan 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, 7 Agustus 2018



Sandy Kurniawan
NIM. 24010314120009

HALAMAN PENGESAHAN

Judul : *Hierarchical Sentence Sentiment Analysis* untuk Ulasan Hotel pada
Website Traveloka Menggunakan Metode Naive Bayes Classifier
Nama : Sandy Kurniawan
NIM : 24010314120009

Telah diujikan pada sidang skripsi tanggal 30 Juli 2018 dan dinyatakan lulus pada tanggal 30 Juli 2018.

Semarang, 7 Agustus 2018

Mengetahui,

Ketua Departemen Ilmu Komputer/ Informatika



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

Panitia Penguji Skripsi

Ketua,

A handwritten signature in black ink, appearing to be 'Priyo Sidik S.', written over a horizontal line.

Priyo Sidik S., S.Si., M.Kom.
NIP. 197007051997021001

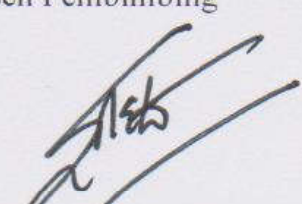
HALAMAN PENGESAHAN

Judul : *Hierarchical Sentence Sentiment Analysis* untuk Ulasan Hotel pada
Website Traveloka Menggunakan Metode Naïve Bayes Classifier
Nama : Sandy Kurniawan
NIM : 24010314120009

Telah diujikan pada sidang skripsi tanggal 30 Juli 2018.

Semarang, 7 Agustus 2018

Dosen Pembimbing



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

ABSTRAK

Traveloka menyediakan ruang bagi penggunanya untuk menuliskan ulasan tentang layanan hotel yang disewakan. Ulasan-ulasan ini sangat berguna bagi pengelola hotel dalam mengetahui tingkat kepuasan pelanggan. *Sentiment analysis* merupakan sarana yang dapat digunakan untuk menganalisis ulasan tersebut, sehingga dapat diketahui apakah ulasan tersebut mengandung opini atau tidak, yang kemudian tingkat kepuasan pelanggan akan diukur berdasarkan jumlah sentimen (positif, negatif) yang terkandung dari opini yang didapatkan. Penelitian ini menggunakan metode *Naïve Bayes classifier* dalam melakukan *hierarchical sentence sentiment analysis* pada ulasan hotel dari *website* Traveloka. Selain itu, penelitian ini menggunakan dua jenis pembobotan kata dalam ekstraksi fitur, yaitu *raw term frequency* dan TF-IDF. *Sentence sentiment analysis* menggunakan *flat classification* untuk ulasan hotel pada *website* Traveloka dilakukan untuk membandingkan hasilnya dengan *hierarchical sentence sentiment analysis* pada ulasan hotel dari *website* Traveloka. Hasil penelitian *sentiment analysis* terhadap ulasan hotel ini menunjukkan bahwa penggunaan *hierarchical classification* dalam *sentiment analysis* lebih baik dibandingkan dengan menggunakan *flat classification*. Hasil perhitungan rata-rata nilai *f-measure* untuk model *flat classification* menghasilkan nilai sebesar 0,7518, sedangkan untuk model *hierarchical classification* menghasilkan nilai sebesar 0,7748. Berdasarkan hasil tersebut, menunjukkan bahwa penggunaan *hierarchical classification* pada *sentiment analysis* meningkatkan kinerja rata-rata model klasifikasi sebesar 0,023. Penggunaan *fitur raw term frequency* pada *flat classification* memberikan nilai *f-measure* yang lebih tinggi daripada penggunaan fitur TF-IDF dengan selisih 0,039. Rata-rata nilai *f-measure flat classification* dengan fitur *raw term frequency* menghasilkan nilai sebesar 0,7518 sedangkan untuk fitur TF-IDF sebesar 0,7123.

Kata Kunci : Traveloka, ulasan, *sentiment analysis*, *hierarchical classification*, *Naïve Bayes classifier*

ABSTRACT

Traveloka provides space for its users to write reviews about their hotel reservation services. These reviews are very useful for the hotel manager in knowing the level of customer satisfaction. Sentiment analysis is a tool that can be used to analyze the reviews, so it can know whether the reviews contain opinion or not, then the level of customer satisfaction will be measured based on the number of sentiments (positive or negative) contained in the opinion. In this research, the Naïve Bayes classifier method was used to perform hierarchical sentence sentiment analysis on the hotel reviews obtained from Traveloka. In addition, this research used two types of term weighting schemes for feature extraction, raw term frequency and TF-IDF. Sentence sentiment analysis using flat classification on hotel reviews from Traveloka were conducted to compare the results with the hierarchical sentence sentiment analysis on hotel reviews from Traveloka. The results of this research about sentiment analysis on hotel reviews indicated that the use of hierarchical classification in sentiment analysis was better than flat classification. The average f-measure value for flat classification model was 0.7518, while the average f-measure value for hierarchical classification model was 0.7748. Based on these results, showed that the use of hierarchical classification in sentiment analysis improved the average performance of the classification model by 0.023. The use of the raw term frequency feature extraction in flat classification provided a higher f-measure value than the use of the TF-IDF feature extraction, with a margin of 0.039. The average value of f-measure for flat classification using raw term frequency feature extraction was 0.7518 while for TF-IDF feature extraction was 0.7123.

Keywords : Traveloka, reviews, *sentiment analysis, hierarchical classification, Naïve Bayes classifier*

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “*Hierarchical Sentence Sentiment Analysis* untuk Ulasan Hotel pada *Website Traveloka Menggunakan Metode Naïve Bayes Classifier*”.

Skripsi ini dibuat dengan tujuan sebagai salah satu syarat untuk memperoleh gelar sarjana komputer pada Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro, Semarang.

Dalam pelaksanaan skripsi serta penyusunan dokumen skripsi ini, penulis menyadari banyak pihak yang membantu sehingga akhirnya dokumen ini dapat diselesaikan. Oleh karena itu, melalui kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Retno Kusumaningrum, S.Si, M.Kom. selaku Kepala Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro, Semarang, sekaligus menjadi Dosen Pembimbing yang telah membimbing dalam penyelesaian skripsi ini.
2. Bapak Helmie Arif Wibawa, S.Si, M.Cs, selaku Koordinator Skripsi Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro, Semarang.
3. Orang tua, keluarga, sahabat dan teman-teman yang telah mendukung, membantu dan memberikan semangat kepada penulis dalam menyelesaikan skripsi ini.
4. Semua pihak yang telah membantu kelancaran dalam penyusunan skripsi, yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan laporan skripsi ini. Oleh karena itu, saran dan kritik yang membangun sangat penulis harapkan. Semoga laporan skripsi ini dapat bermanfaat bagi semua pihak.

Semarang, 7 Agustus 2018

Sandy Kurniawan

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
DAFTAR LAMPIRAN	xiv
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	5
BAB II TINJAUAN PUSTAKA	6
2.1. Perkembangan Penelitian Mengenai <i>Sentiment Analysis</i> Dokumen Berbahasa Indonesia.....	6
2.2. <i>Preprocessing</i>	7
2.2.1. <i>Case Folding</i>	7
2.2.2. Tokenisasi.....	7
2.2.3. <i>Filtering</i>	7
2.2.4. <i>Stemming</i>	8
2.3. <i>Term Weighting</i>	15
2.4. <i>K-Fold Cross-Validation</i>	16
2.5. <i>Naïve Bayes classifier</i>	17
2.6. Evaluasi.....	19
2.7. Pengembangan Perangkat Lunak.....	20
BAB III METODOLOGI PENELITIAN	22
3.1. Pengumpulan Dataset	23

3.2. <i>Preprocessing</i>	24
3.2.1. Case Folding.....	24
3.2.2. Tokenisasi.....	25
3.2.3. <i>Filtering</i>	26
3.2.4. <i>Stemming</i>	27
3.3. Ekstraksi Fitur.....	32
3.3.1. Pembuatan Kamus Kata	33
3.3.2. Pembuatan Fitur <i>Raw term frequency</i>	34
3.3.3. Pembuatan Fitur <i>Term Frequency – Inverse Document Frequency</i>	35
3.4. <i>K-Fold Cross Validation</i>	38
3.5. Pelatihan	39
3.5.1. Pelatihan <i>Flat Classification</i>	39
3.5.2. Pelatihan <i>Hierarchical Classification</i>	44
3.6. Pengujian	47
3.6.1. Pengujian <i>Flat Classification</i>	48
3.6.2. Pengujian <i>Hierarchical Classification</i>	49
3.7. Evaluasi.....	54
3.8. Proses <i>Data crawling</i>	56
3.9. <i>Preprocessing</i> Ulasan Baru	56
3.10. <i>Sentiment analysis</i>	56
3.11. Analisis dan Desain Sistem.....	57
3.11.1. Analisis Perangkat Lunak	58
3.11.2. Perancangan Perangkat Lunak.....	59
BAB IV HASIL DAN ANALISA.....	66
4.1. Hasil Pengembangan Perangkat Lunak	66
4.1.1. Lingkungan Implementasi Perangkat Lunak.....	66
4.1.2. Implementasi Antarmuka	66
4.2. Skenario Pengujian Perangkat Lunak.....	71
4.2.1. Skenario Pengujian Fungsional Perangkat Lunak.....	71
4.2.2. Skenario Pengujian Kinerja Perangkat Lunak	71
4.3. Hasil dan Analisa Perangkat Lunak.....	73
4.3.1. Hasil dan Analisa Pengujian Fungsional Perangkat Lunak	73
4.3.2. Hasil dan Analisa Pengujian Kinerja Perangkat Lunak	73

BAB V PENUTUP	82
5.1. Kesimpulan	82
5.2. Saran	83
DAFTAR PUSTAKA.....	84

DAFTAR GAMBAR

Gambar 2.1 <i>Confusion Matrix</i> (Doreswamy, 2011)	19
Gambar 2.2 <i>Software Process Model Waterfall</i> (Sommerville, 2011).....	20
Gambar 3.1 Gambaran Umum Penelitian	22
Gambar 3.2 <i>Flowchart</i> Tokenisasi	25
Gambar 3.3 <i>Flowchart Filtering</i>	27
Gambar 3.4 <i>Flowchart Stemming</i>	28
Gambar 3.5 <i>Flowchart Stemmer</i> Sastrawi (Bashri, 2017).....	28
Gambar 3.6 <i>Flowchart</i> Sub-proses <i>StemmingPlural</i> (Bashri, 2017).....	29
Gambar 3.7 <i>Flowchart</i> Sub-proses <i>StemmingSingular</i> (Bashri, 2017)	30
Gambar 3.8 <i>Flowchart</i> Sub-proses <i>LoopPengembalianAkhiran</i> (Bashri, 2017).....	32
Gambar 3.9 Alur Ekstraksi Fitur	33
Gambar 3.10 <i>Flowchart</i> Pembuatan Kamus Kata (<i>Vocabulary</i>).....	34
Gambar 3.11 <i>Flowchart</i> Pembuatan Fitur <i>Raw term frequency</i>	35
Gambar 3.12 <i>Flowchart</i> Pembuatan Fitur TF-IDF.....	36
Gambar 3.13 Ilustrasi Pembagian Dataset untuk <i>10-Fold Cross Validation</i>	38
Gambar 3.14 <i>Flowchart</i> Pelatihan <i>Flat Classification</i>	39
Gambar 3.15 <i>Flowchart</i> Pelatihan <i>Naïve Bayes Classification</i>	40
Gambar 3.16 <i>Hierarchical Classification</i>	45
Gambar 3.17 <i>Flowchart</i> Pelatihan <i>Hierarchical Classification</i>	45
Gambar 3.18 <i>Flowchart</i> Pelatihan <i>Hierarchical Classification Level 1</i>	46
Gambar 3.19 <i>Flowchart</i> Pelatihan <i>Hierarchical Classification Level 2</i>	47
Gambar 3.20 <i>Flowchart</i> Pengujian <i>Flat Classification</i>	48
Gambar 3.21 <i>Flowchart</i> Pengujian <i>Hierarchical Classification</i>	50
Gambar 3.22 <i>Flowchart</i> Pengujian <i>Hierarchical Classification Level 1</i>	51
Gambar 3.23 <i>Flowchart</i> Pengujian <i>Hierarchical Classification Level 2</i>	52
Gambar 3.24 <i>Flowchart</i> Pengujian <i>Hierarchical Classification</i>	53
Gambar 3.25 <i>Flowchart Real Time Sentiment Analysis</i>	57
Gambar 3.26 <i>Data Context Diagram</i> (DCD) HSSA	59
Gambar 3.27 <i>Data Flow Diagram Level 1</i> HSSA.....	60
Gambar 3.28 Antarmuka Halaman Awal	62
Gambar 3.29 Antarmuka Halaman <i>Input Parameter Live crawling</i>	62
Gambar 3.30 Antarmuka Halaman Hasil <i>Real Time Sentiment Analysis</i> Tingkat Dokumen	63
Gambar 3.31 Antarmuka Halaman Detail Klasifikasi Tingkat Kalimat	63
Gambar 3.32 Antarmuka Halaman Dataset Ulasan.....	64
Gambar 3.33 Antarmuka Halaman <i>Preprocessing</i>	64
Gambar 3.34 Antarmuka Halaman Hasil Pengujian Model.....	65
Gambar 4.1 Implementasi Antarmuka Halaman Awal	67
Gambar 4.2 Implementasi Antarmuka Halaman <i>Input Parameter Live crawling</i>	67

Gambar 4.3 Implementasi Antarmuka Halaman Hasil <i>Real Time Sentiment analysis</i> Tingkat Dokumen.....	68
Gambar 4.4 Implementasi Antarmuka Halaman Detail Klasifikasi Tingkat Kalimat.....	69
Gambar 4.5 Implementasi Antarmuka Halaman Dataset Ulasan	69
Gambar 4.6 Implementasi Antarmuka Halaman <i>Preprocessing</i>	70
Gambar 4.7 Implementasi Antarmuka Halaman Hasil Pengujian Model	70
Gambar 4.8 Presentase Dataset Ulasan	72
Gambar 4.9 Grafik Nilai <i>F-measure</i> Skenario 1	75
Gambar 4.10 Hasil Kinerja Model <i>Classifier Level 1</i>	77
Gambar 4.11 Hasil Kinerja Model <i>Classifier Level 2</i>	79
Gambar 4.12 Hasil Kinerja <i>Hierarchical Classification</i>	80
Gambar 4.13 Hasil Kinerja <i>Flat dan Hierarchical Classification</i>	81

DAFTAR TABEL

Tabel 2.1 Penelitian Lain Terkait <i>Sentiment Analysis</i> Dengan Bahasa Indonesia	6
Tabel 2.2 Aturan Pemenggalan Awalan Algoritma Nazief-Adriani	10
Tabel 2.3 Modifikasi Dan Tambahan Aturan Oleh Algoritma <i>Confix Stripping</i>	12
Tabel 2.4 Daftar <i>Rule Precedence</i>	12
Tabel 2.5 Modifikasi Aturan Oleh Algoritma <i>Enhanced Confix Stripping</i>	13
Tabel 2.6 Modifikasi Aturan Pemenggalan Kata Algoritma <i>Enhanced Confix Stripping Stemmer</i>	14
Tabel 2.7 Modifikasi Aturan Yang Digunakan Pada <i>Stemmer</i> Sastrawi.....	15
Tabel 3.1 Contoh Dataset Ulasan Sebelum Dilakukan <i>Case Folding</i>	24
Tabel 3.2 Contoh Dataset Ulasan Setelah Dilakukan <i>Case Folding</i>	24
Tabel 3.3 Contoh Dataset Ulasan Sebelum Dilakukan Proses Tokenisasi.....	25
Tabel 3.4 Contoh Dataset Ulasan Setelah Dilakukan Proses Tokenisasi	25
Tabel 3.5 Contoh Dataset Ulasan Sebelum Dilakukan <i>Filtering</i>	26
Tabel 3.6 Contoh Dataset Ulasan Setelah Dilakukan <i>Filtering</i>	26
Tabel 3.7 Contoh Dataset Ulasan Sebelum Proses <i>Stemming</i>	31
Tabel 3.8 Contoh Dataset Ulasan Setelah Proses <i>Stemming</i>	32
Tabel 3.9 Contoh Dataset Ulasan Hasil <i>Preprocessing</i>	34
Tabel 3.10 Contoh Hasil Kamus Kata Yang Didapatkan	34
Tabel 3.11 Contoh Representasi <i>Vector</i> Fitur Ulasan Dengan <i>Raw Term Frequency</i>	35
Tabel 3.12 Perhitungan TF Dan Df	36
Tabel 3.13 Perhitungan IDF	37
Tabel 3.14 Perhitungan TF-IDF	37
Tabel 3.15 Contoh Representasi Vektor Ulasan Dengan TF-IDF.....	37
Tabel 3.16 Dataset Pelatihan	41
Tabel 3.17 Hasil <i>Preprocessing</i> Dataset Latihan	41
Tabel 3.18 Representasi Dokumen Dalam Bentuk Fitur <i>Vector</i>	41
Tabel 3.19 Nilai <i>NxiNeutral</i> Dan <i>NNeutral</i> Dari Dokumen Untuk Kelas Netral	43
Tabel 3.20 Nilai <i>NxiPositif</i> Dan <i>NPositif</i> Dari Dokumen Untuk Kelas Positif.....	43
Tabel 3.21 Nilai <i>NxiNegatif</i> Dan <i>NNegatif</i> Dari Dokumen Untuk Kelas Negatif....	43
Tabel 3.22 Representasi Fitur <i>Vector</i> Data Baru.....	49
Tabel 3.23 Contoh Hasil Klasifikasi Dokumen.....	54
Tabel 3.24 Contoh <i>Confusion Matrix</i>	54
Tabel 3.25 Kebutuhan Fungsional.....	58
Tabel 3.26 Kebutuhan Non-Fungsional	58
Tabel 4.1 Butir Pengujian Fungsional Perangkat Lunak	71
Tabel 4.2 Hasil Skenario 1	74
Tabel 4.3 Hasil <i>10-Fold Cross Validation Classifier Level 1</i>	76
Tabel 4.4 Hasil <i>10-Fold Cross Validation Classifier Level 2</i>	77
Tabel 4.5 Hasil <i>10-Fold Cross Validation Hierarchical Classification</i>	79
Tabel 4.6 Perbandingan <i>Flat Classification</i> Dan <i>Hierarchical Classification</i>	80

DAFTAR LAMPIRAN

Lampiran 1. Hasil Pengujian Fungsional Perangkat Lunak	88
Lampiran 2. Hasil Skenario 1	89
Lampiran 3. Hasil Skenario 2	91

BAB I

PENDAHULUAN

Bab pendahuluan membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan dalam pembuatan skripsi mengenai *hierarchical sentence sentiment analysis* untuk ulasan hotel pada *Website Traveloka* Menggunakan Metode *Naïve Bayes classifier*.

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini sangatlah pesat, hal ini dikarenakan teknologi informasi merupakan jawaban dari permasalahan dalam berbagai bidang di kehidupan manusia. Salah satu bidang yang sangat berkembang berkat bantuan teknologi informasi adalah bidang bisnis, sebagai contoh *electronic-commerce (e-commerce)*. Konsumen dimudahkan dalam mendapatkan barang atau jasa yang dibutuhkan dengan adanya *e-commerce* ini (Syafik & Tanamal, 2017). *E-commerce* hadir dengan berbagai bentuk salah satunya adalah Traveloka, sebuah *e-commerce Online Travel Agents (OTA)* yang bergerak dalam bidang pemesanan hotel, tiket pesawat, tiket kereta api dan lain sebagainya. Selain menyediakan layanan pemesanan hotel, pesawat serta kereta api, Traveloka juga menyediakan fitur ulasan mengenai pelayanan yang diberikan oleh jasa yang telah dipesan melalui Traveloka, sehingga dapat membantu pengguna dalam memilih hotel, pesawat maupun kereta api yang direkomendasikan. Ulasan tersebut juga dapat dianalisa untuk membantu pengelola jasa (hotel, pesawat dan kereta api) menilai bagaimana tanggapan konsumen terhadap jasa yang diberikan. Ulasan positif tentu akan meningkatkan *rating* serta popularitas dari jasa tersebut, namun dengan adanya ulasan negatif dapat menjadi bahan evaluasi pengelola jasa untuk memperbaiki jasa yang diberikan. Akan tetapi, dengan banyaknya jumlah ulasan yang diberikan mempersulit pengelola jasa dalam melakukan analisa ulasan tersebut. Oleh karena itu diperlukan *sentiment analysis* untuk mengolah data serta menganalisa ulasan yang ada.

Sentiment analysis atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini (Rozi, et al., 2012).

Sentiment analysis dapat digunakan untuk mengklasifikasikan ulasan ke dalam sentimen tertentu dengan memperhatikan faktor pada objek sentimennya. Penelitian *sentiment analysis* ini menggunakan ulasan hotel sebagai objek penelitian, sehingga memperhatikan berbagai faktor seperti pelayanan, keadaan dan harga sewa hotel dalam penentuan kelas setiap ulasannya. Ulasan akan diklasifikasikan menjadi 3 kelas, yaitu ulasan netral, ulasan positif dan ulasan negatif.

Penelitian mengenai *sentiment analysis* dalam Bahasa Indonesia pernah dilakukan oleh Zulfa dan Winarko (2017) dengan menggunakan metode *Deep Belief Network* (DBN) dengan objek penelitian *tweet* dari media sosial Twitter. Berdasarkan penelitian tersebut, dijelaskan bahwa metode DBN dengan menggunakan *Bag of Word* (BoW) sebagai fitur ekstraksinya tidak memberikan akurasi yang lebih baik dibandingkan dengan metode *Naïve Bayes classifier* dan *Support Vector Machine* (SVM) dengan ekstraksi fitur yang sama. Selain itu juga ada penelitian mengenai *sentiment analysis* dengan objek penelitian ringkasan *review* film dengan menggunakan metode *Information Gain* dan *K-Nearest Neighbor* yang dilakukan oleh Pristiyanti, et al (2018). Penelitian tersebut menyimpulkan bahwa metode *Information Gain* tidak dapat membedakan antara *term sentiment analysis* dengan kata bukan *sentiment analysis*, oleh karena itu diperlukan metode yang dapat membedakan antara *term sentiment analysis* atau bukan. Penelitian mengenai *sentiment analysis* yang lain dengan menggunakan metode *Naïve Bayes classifier* juga dilakukan oleh Lestari, et al (2017). Penelitian tersebut menyimpulkan bahwa metode *Naïve Bayes classifier* dengan pembobotan emoji dapat diterapkan pada penelitian *sentiment analysis* dan pembobotan emoji dapat meningkatkan akurasi dari sistem yang dibuat (Lestari, et al., 2017).

Oleh karena itu, pada penelitian ini akan diterapkan metode *Naïve Bayes classifier* untuk melakukan *sentiment analysis*. Penggunaan metode *Naïve Bayes classifier* didasarkan pada kecepatan komputasi dan tingginya akurasi yang dihasilkan dalam semua domain yang berkaitan dengan *sentiment analysis* (Sharada & Krishna, 2017). Aggarwal & Zhai (2012) menyebutkan bahwa metode *Naïve Bayes classifier* cocok untuk digunakan pada *hierarchical classification* dengan dataset latih yang digunakan disusun berdasarkan taksonomi dari topik yang digunakan. Penelitian mengenai *hierarchical classification* telah dilakukan oleh Silla & Freitas (2011), dimana Silla & Freitas telah melakukan survei mengenai *hierarchical classification*

dalam aplikasi berbagai domain seperti *text categorization*, *music genre classification* dan *image classification* dalam *document level*. Dalam penelitian tersebut Silla & Freitas merangkum berbagai penelitian mengenai *hierarchical classification* serta perbandingan hasil kinerja *hierarchical classification* dengan *flat classification*. Sebagian besar hasil perbandingan tersebut, menunjukkan bahwa *hierarchical classification* mendapatkan kinerja prediksi yang lebih baik dibandingkan *flat classification*.

Di sisi lain, Liu (2012) menyebutkan bahwa dalam *sentiment analysis*, terdapat 3 level yang dapat digunakan dalam penelitian yaitu *document level*, *sentence level* dan *aspect level*. *Document level sentiment analysis* mengklasifikasikan sebuah opini secara keseluruhan, apakah mengandung sentimen positif atau negatif. *Sentence level sentiment analysis* mengklasifikasikan setiap kalimat penyusun dalam suatu opini apakah mengandung sentimen positif, negatif atau tidak mengandung sentimen sama sekali (netral). *Aspect level sentiment analysis* melakukan analisis sentimen terhadap aspek tertentu yang ditentukan terlebih dahulu, apakah mengandung sentimen positif atau negatif. Penelitian ini menggunakan *sentence level sentiment analysis* untuk mengetahui kinerja *hierarchical classification* terhadap *sentence level classification*, apakah akan berpengaruh seperti pada *document level classification* atau tidak. Liu (2012) mendefinisikan permasalahan pada *sentence level sentiment analysis* adalah untuk menentukan apakah suatu kalimat menyatakan opini positif, negatif atau netral (atau tidak mengandung opini). Permasalahan ini dapat diselesaikan dengan menggunakan dua tahap/langkah klasifikasi yang terpisah. Langkah pertama adalah untuk mengklasifikasikan apakah suatu kalimat menunjukkan sebuah opini atau tidak. Langkah ini juga disebut dengan *subjectivity classification*, yang mana menentukan apakah kalimat mengandung informasi subjektif atau informasi objektif (faktual). Kemudian pada langkah kedua akan mengklasifikasikan kalimat terklasifikasi opini menjadi kelas positif atau negatif.

Oleh karena itu, pada penelitian ini akan dikaji mengenai penerapan *hierarchical sentence sentiment analysis* pada ulasan hotel dari *website* Traveloka menggunakan metode *Naïve Bayes classifier*.

1.2 Rumusan Masalah

Berdasarkan pada latar belakang dapat dirumuskan permasalahan yang dapat diambil yaitu bagaimana menerapkan *hierarchical sentence sentiment analysis* menggunakan metode *Naïve Bayes classifier* terhadap ulasan pada *website* Traveloka dan membandingkan hasilnya dengan *flat sentence sentiment analysis* dengan menggunakan metode *Naïve Bayes classifier* terhadap ulasan pada *website* Traveloka.

1.3 Tujuan dan Manfaat

Tujuan umum dari penelitian skripsi ini adalah melakukan *hierarchical sentence sentiment analysis* terhadap ulasan pada Traveloka menggunakan metode *Naïve Bayes classifier*. Adapun tujuan khusus dari penelitian skripsi ini, antara lain:

1. Mengetahui kinerja *flat sentiment analysis* dengan metode *Naïve Bayes classifier*.
2. Mengetahui kinerja *hierarchical sentiment analysis* dengan metode *Naïve Bayes classifier*.
3. Membandingkan kinerja *flat sentiment analysis* dengan *hierarchical sentiment analysis* dengan metode *Naïve Bayes classifier*.

Sedangkan manfaat yang diharapkan dari penelitian skripsi ini adalah menghasilkan *sentiment analysis* yang dapat dimanfaatkan untuk mengetahui sentimen masyarakat terhadap suatu hotel bagi pengunjung maupun pengelola hotel melalui ulasan di Traveloka.

1.4 Ruang Lingkup

Ruang lingkup dalam menerapkan *sentiment analysis* menggunakan metode *Naïve Bayes classifier* adalah sebagai berikut:

1. Objek penelitian yang digunakan pada skripsi ini adalah ulasan hotel yang disediakan oleh Traveloka.
2. Data ulasan yang digunakan diperoleh berdasarkan urutan waktu terbaru dari setiap ulasan hotel yang disediakan oleh Traveloka hingga bulan Maret tahun 2018.
3. Data diperoleh dengan cara *crawling* ulasan pada Traveloka sebanyak 1720 data ulasan berupa data dengan jumlah seimbang untuk beberapa hotel yang terdiri dari 860 data non-sentimen, 430 data sentimen positif dan 430 data sentimen negatif.

1.5 Sistematika Penulisan

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

BAB I PENDAHULUAN

Bab pendahuluan membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan dalam penyusunan skripsi ini.

BAB II LANDASAN TEORI

Bab ini membahas mengenai kajian pustaka yang berhubungan dengan skripsi sebagai landasan untuk merumuskan dan menganalisa permasalahan pada skripsi. Kajian pustaka yang digunakan meliputi *sentiment analysis*, *preprocessing*, *term weighting*, *K-fold cross-validation*, *Naïve Bayes classifier*, evaluasi dan pengembangan perangkat lunak.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan yang dilakukan dalam penelitian skripsi. Tahapan tersebut meliputi pengumpulan data, proporsi dataset, *preprocessing*, ekstraksi fitur, *k-fold cross validation*, pelatihan-pengujian, evaluasi dan pengembangan aplikasi *sentiment analysis*.

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.