

**METODE "GLOBAL VECTOR" (GLOVE) UNTUK REPRESENTASI
VEKTOR KATA PADA ULASAN HOTEL DI INDONESIA**

SKRIPSI



disusun oleh

Lutfi Dian Mulyanto

17.11.0953

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

**METODE "GLOBAL VECTOR" (GLOVE) UNTUK REPRESENTASI
VEKTOR KATA PADA ULASAN HOTEL DI INDONESIA**

SKRIPSI

Untuk memenuhi sebagian persyaratan
Mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Lutfi Dian Mulyanto

17.11.0953

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

PERSETUJUAN

SKRIPSI

**METODE "GLOBAL VECTOR" (GLOVE) UNTUK REPRESENTASI
VEKTOR KATA PADA ULASAN HOTEL DI INDONESIA**

yang dipersiapkan dan disusun oleh

Lutfi Dian Mulyanto

17.11.0953

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 September 2020

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom.

NIK. 190302108

PENGESAHAN**SKRIPSI****METODE "GLOBAL VECTOR" (GLOVE) UNTUK REPRESENTASI
VEKTOR KATA PADA ULASAN HOTEL DI INDONESIA**

yang dipersiapkan dan disusun oleh

Lutfi Dian Mulyanto

17.11.0953

telah dipertahankan di depan Dewan Penguji
pada tanggal 21 September 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mardhiya Hayaty, S.T., M.Kom.
NIK. 190302108

Ike Verawati, M.Kom.
NIK. 190302237

Wiwi Widayani, M.Kom.
NIK. 190302227

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 September 2020

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan di bawah ini menyatakan bahwa skripsi ini merupakan karya saya sendiri (ASLI) dan isi pada skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan masalah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 26 September 2020



Lutfi Dian Mulyanto

NIM 17.11.0953

MOTTO

“The Biggest Risk Is Not Taking Any Risk”

-Mark Zuckerberg

"Stay Hungry, Stay Foolish."

-Steve Jobs

“Sebaik-baiknya manusia adalah yang bermanfaat bagi orang lain”

-HR.Ahmad

“Don't dream your life, live your dream”

-Mesut Ozil



PERSEMBAHAN

Alhamdulillah dengan kerja keras serta doa, skripsi ini dapat diselesaikan dengan baik. Segala puji dan syukur bagi Allah SWT yang tiada henti memberikan keberkahan. Dengan ini saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung, yaitu untuk :

1. Kedua orang tua, dan kakak saya, yang selalu mendoakan, selalu menyemangati dan memberikan motivasi tiada henti.
2. Dosen pembimbing saya Ibu Mardhiya Hayaty, S.T., M.Kom., yang telah membimbing saya dari awal sampai akhir pembuatan skripsi.
3. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
4. Teman-teman satu kloter penelitian yang memberikan dukungan dan motivasi, serta menemani saya untuk menyelesaikan skripsi ini.
5. Teman-teman kelas 17-IF-01 yang telah menemani dan selalu memberikan semangat untuk menyelesaikan skripsi ini.
6. Teman-teman SRAWUNG (Aldo, Dhimsky, Andrian, Fariz, Rizky, Bapok, Irfan, Anjas dan Vina) yang hingga saat ini menjadi teman disaat butuh liburan yang selalu memberikan dukungan sampai saat ini.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya dan shawalat serta salam juga tidak lupa penulis panjatkan kepada junjungan kita Nabi Muhammad SAW yang telah memberikan teladan mulia dalam menuntun umatnya sehingga penulis dapat menyelesaikan skripsi ini.

Skripsi yang berjudul **“METODE”GLOBAL VECTOR”(GLOVE) UNTUK REPRESENTASI VEKTOR KATA PADA ULASAN HOTEL DI INDONESIA“** ini disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas AMIKOM Yogyakarta.

Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada:

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Bapak Sudarmawan, M.T. selaku ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Mardhiya Hayaty, S.T., M.Kom. selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, nasehat serta waktunya selama penulisan skripsi ini.

5. Ibu Ike Verawati, M.Kom. dan Ibu Wiwi Widayani, M.Kom. selaku dosen penguji. Terimakasih atas saran yang diberikan selama pengujian untuk memperbaiki penelitian menjadi lebih baik lagi.

Penulis menyadari skripsi ini masih ada kekurangan. Maka, penulis menerima kritik dan saran yang membangun serta teguran dari semua pihak. Penulis menerima dengan lapang dada untuk kesempurnaan untuk kesempurnaan karya selanjutnya. Semoga skripsi yang sederhana ini bisa bermanfaat. Khususnya bagi penulis dan pembaca yang budiman pada umumnya. Apabila terdapat kesalahan semoga Allah SWT melimpahkan magfirah-Nya. *Aamiin yaa Kholi.*

Yogyakarta, 26 September 2020

Lutfi Dian Mulyanto

DAFTAR ISI

JUDUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN.....	iv
MOTTO	v
PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Maksud dan Tujuan Penelitian	3
1.5 Manfaat Penelitian.....	3
1.6 Metode Penelitian.....	3
1.6.1 Pengumpulan Data	3
1.6.1.1 Manual Labeling	4
1.6.1.2 Preprocessing Data	4
1.6.1.3 Ekstraksi Fitur	5
1.6.1.4 Implementasi Algoritma Klasifikasi.....	5

1.6.1.5	Evaluasi.....	5
1.7	Sistematika Penulisan.....	6
BAB I	PENDAHULUAN	6
BAB II	LANDASAN TEORI.....	6
BAB III	METODOLOGI PENELITIAN	6
BAB IV	HASIL DAN PEMBAHASAN	6
BAB V	PENUTUP.....	6
BAB II	LANDASAN TEORI	7
2.1	Tinjauan Pustaka.....	7
2.2	Dasar Teori.....	10
2.2.1	<i>Natural Language Processing</i>	10
2.2.2	<i>Deep Learning</i>	10
2.2.3	<i>Data Mining</i>	11
2.2.4	<i>Text Mining</i>	11
2.2.5	Analisis Sentimen.....	14
2.2.6	<i>Review</i>	15
2.2.7	<i>Word Embedding</i>	16
2.2.8	<i>Vector Space Model</i>	19
2.2.9	<i>Indexing</i>	22
2.2.10	<i>Global Vector (GloVe)</i>	22
2.2.11	<i>Preprocessing Data</i>	25
2.2.12	<i>Long Short Term Memory</i>	26
2.2.13	Optimasi Adam	29
2.2.14	Batch Size dan Epoch	30
2.2.15	<i>Confusion Matrix</i>	30
BAB III	METODELOGI PENELITIAN.....	33

3.1	Tahapan penelitian	33
3.2	Alat Penelitian	33
3.2.1	Perangkat Keras (<i>Hardware</i>)	33
3.2.2	Perangkat Lunak (<i>Software</i>)	34
3.3	Pengumpulan Data.....	34
3.4	Manual Labeling	34
3.5	<i>Preprocessing</i> Data	35
3.6	Data Training dan Data Testing	36
3.7	Ekstraksi Fitur	36
3.8	Implementasi Algoritma Klasifikasi.....	37
3.9	Evaluasi.....	37
BAB IV IMPLEMENTASI DAN PEMBAHASAN		39
4.1	Pengumpulan Data.....	39
4.2	Dataset.....	40
4.3	<i>Preprocessing</i> Data	41
4.4	Ekstraksi Fitur	48
4.5	Implementasi Algoritma LSTM	53
4.6	Hasil Training Data	57
4.7	Pengukuran Algoritma.....	59
4.8	Pembahasan.....	63
BAB V PENUTUP.....		64
5.1	Kesimpulan.....	64
5.2	Saran	64
DAFTAR PUSTAKA.....		65

DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian.....	9
Tabel 2. 2 Confusion matrix	31
Tabel 4. 1 Contoh ulasan hotel positif dan negatif.....	41
Tabel 4. 2 Hasil Data Cleaning	43
Tabel 4. 3 Hasil Case Folding.....	44
Tabel 4. 4 Hasil Tokenizing	46
Tabel 4. 5 Hasil Stopword Removal	47
Tabel 4. 7 Hasil Pemberian Indeks Pada Ulasan.....	49
Tabel 4. 8 Hasil Pad_Sequence.....	50
Tabel 4. 9 Hasil Word Embedding Layer	51
Tabel 4. 10 Confusion Matrix non Word Embedding.....	61
Tabel 4. 11 Confusion Matrix Word Embedding	61
Tabel 4. 12 Evaluasi Perbandingan Confusion Matrix	62

DAFTAR GAMBAR

Gambar 2. 1 Alur proses text mining.....	13
Gambar 2. 2 Ilustrasi mapping sentimen word embedding.....	17
Gambar 2. 3 Word Embedding	18
Gambar 2. 4 Representasi Model Ruang Vektor	20
Gambar 2. 5 Representasi nilai kata dalam dokumen.....	21
Gambar 2. 6 Arsitektur LSTM.....	27
Gambar 3. 1 Diagram Alir Tahapan Penelitian.....	33
Gambar 4. 1 Scraping data menggunakan WebHarvy.....	39
Gambar 4. 2 Dataset.....	40
Gambar 4. 3 Script Cleaning Tag HTML	42
Gambar 4. 4 Script Data Cleaning	42
Gambar 4. 5 Script Case Folding.....	43
Gambar 4. 6 Script Tokenizing.....	45
Gambar 4. 7 Script Stopword Removal	46
Gambar 4. 9 Split Dataset	47
Gambar 4. 10 Script pre-trained Glove.....	48
Gambar 4. 11 Script Proses Word Embedding Dari Dataset.....	49
Gambar 4. 12 Visualisasi Word Embedding dataset.....	53
Gambar 4. 13 Script Train LSTM.....	54
Gambar 4. 14 Total Parameter.....	55
Gambar 4. 15 Architecture Model LSTM.....	56
Gambar 4. 16 Script Fit <i>Network LSTM</i>	57
Gambar 4. 17 Hasil Training dan Validasi non Word Embedding.....	58
Gambar 4. 18 Hasil Training dan Validasi Word Embedding	58
Gambar 4. 19 Script Evaluasi Model.....	59
Gambar 4. 20 Script Prediksi Model.....	60
Gambar 4. 21 Script Confusion Matrix.....	60

INTISARI

Perkembangan teknologi informasi membuat produksi data meningkat drastis. Kita bisa mendapatkan banyak data dari internet, termasuk data ulasan tentang suatu produk atau layanan. Semakin banyak data didapat system dibutuhkan untuk mengolahnya. Sentimen analisis adalah pemrosesan teks dari *Natural Language Processing* (NLP) yang dapat membantu untuk melihat kualitas layanan yang ditawarkan, termasuk layanan hotel. Penelitian ini menggunakan data ulasan hotel untuk melaksanakannya analisis sentimen yang diperoleh dari situ web Traveloka.

Data diklasifikasikan menggunakan *Long Short Term Memory* (LSTM). Untuk mendapatkan hasil yang lebih baik, penulis menggunakan *word embedding* kata untuk mengubah kata menjadi vektor. Penelitian ini bertujuan untuk membandingkan kinerja sentimen analisis menggunakan *word embedding* dengan tidak menggunakan *word embedding*.

Metode *word embedding* yang digunakan *Global Vector* (Glove). Dari percobaan yang dilakukan, hasilnya terlihat bahwa metode *word embedding Global Vector* (Glove) memiliki keakuratan tinggi 98%, presisi 97%, *recall* 98% dan *f1-score* 98%. Sedangkan tanpa menggunakan *word embedding* memiliki keakuratan 97%, presisi 95%, *recall* 98%, dan *f1-score* 97%. Akan tetapi untuk kecepatan epoch lebih tinggi tidak menggunakan *word embedding* dengan akurasi terbaik di epoch 6 dan untuk menggunakan *word embedding* di epoch 12. Sehingga disimpulkan bahwa sentimen analisis menggunakan *word embedding* lebih khususnya metode *Global Vector* (Glove) tidak ada perbedaan secara signifikan dengan tanpa menggunakan *word embedding* hanya berbeda 1% tingkat akurasi.

Kata-kunci: Analisis Sentimen, *Word Embedding*, *Glove*, *LSTM*

ABSTRACT

The development of information technology has made data production increase dramatically. We can get a lot of data from the internet, including data reviews about a product or service. The more data obtained the system needed to process it. Sentiment analysis is text processing from Natural Language Processing (NLP) which can help to see the quality of services offered, including hotel services. This study uses hotel review data to carry out a sentiment analysis obtained from the Traveloka website.

Data is classified using Long Short Term Memory (LSTM). To get better results, the author uses word embedding to convert words into vectors. This study aims to compare the performance of the sentiment analysis using word embedding and not using word embedding.

The word embedding method used by Global Vector (Glove). From the experiments conducted, the results show that the word embedding Global Vector (Glove) method has a high accuracy of 98%, 97% precision, 98% recall and 98% f1-score. Meanwhile, without using word embedding, it has an accuracy of 97%, a precision of 95%, a recall of 98%, and f1-score of 97%. However, for higher epoch speeds, don't use word embedding with the best accuracy in epoch 6 and to use word embedding in epoch 12. So it can be concluded that the sentiment analysis using word embedding, especially the Global Vector (Glove) method, has no significant difference without using word embedding, with only 1% difference in accuracy.

Keywords: *Sentiment Analysis, Word Embeddings, Glove, LSTM*