



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

**PERBANDINGAN SENTIMEN ANALISIS TERHADAP
BRAND INDOMIE MENGGUNAKAN NAÏVE BAYES &
LONG SHORT TERM MEMORY (LSTM)**

SKRIPSI



Diajukan Guna Memenuhi Persyaratan Memperoleh
Gelar Sarjana Komputer (S.Kom.)

Elisabeth

14110310073

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2018**

PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi ini adalah karya ilmiah saya sendiri, bukan plagiat dari karya ilmiah yang ditulis oleh orang lain atau lembaga lain, dan semua karya ilmiah orang lain atau lembaga lain yang dirujuk dalam skripsi ini telah disebutkan sumber kutipannya serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan tindak kecurangan / penyimpangan, baik dalam hal pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah skripsi yang telah saya tempuh dan status keserjanaan strata satu yang sudah diterima akan dicabut.

Tangerang, 16 Juli 2018



HALAMAN PENGESAHAN

Skripsi dengan judul

“Perbandingan Sentimen Analisis terhadap *Brand* Indomie Menggunakan
Naïve Bayes & Long Short Term Memory (LSTM)”

oleh

Elisabeth

telah diujikan pada hari Senin, 30 Juli 2018,

pukul 09.30 s.d. 10.30 dan dinyatakan lulus

dengan susunan penguji sebagai berikut.

Ketua Sidang



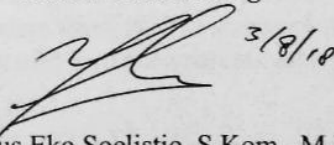
Wira Munggana, S.Si., M.Sc.

Penguji



Ririn Ikana Desanti, S.Kom., M.Kom.

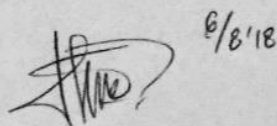
Dosen Pembimbing



Yustinus Eko Soelistio, S.Kom., M.M.

Disahkan oleh

Ketua Program Studi Sistem Informasi - UMN



Ririn Ikana Desanti, S.Kom., M.Kom.

PERBANDINGAN SENTIMEN ANALISIS TERHADAP *BRAND* INDOMIE MENGUNAKAN *NAÏVE BAYES* & *LONG SHORT TERM MEMORY* (LSTM)

ABSTRAK

Oleh: Elisabeth

Sentimen analisis ataupun *opinion mining* merupakan ranah studi yang menganalisa opini, sentimen, evaluasi, sikap, dan emosi penggunaannya dalam bentuk teks tertulis (Liu B. , 2012). Selain diimplementasikan dalam ranah bisnis, sentimen analisis juga kerap kali diperbincangkan dalam ranah penelitian (Vinodhini, 2012). Penelitian sebelumnya terkait klasifikasi sentimen *tweets* dengan menggunakan *Naïve Bayes* (Tunggawan, 2016). Namun demikian, *Naïve Bayes* merupakan metode yang mengasumsikan setiap kata yang muncul secara independen (Ginosar, 2012) dan lebih tepat diimplementasikan pada *datasets* dengan variasi data yang minim (Pang, 2002). Sementara itu, hubungan antar *tweets* cenderung bersifat dependen dan memiliki variasi data yang luas dengan jumlah yang banyak (Jiang, 2011).

Untuk itu, diperlukan eksplorasi lebih lanjut atas pernyataan penelitian sebelumnya (Jiang, 2011). Salah satu upaya eksplorasi yang dapat dilakukan yakni dengan merancang model yang mengasumsikan data secara independen, yakni *Naïve Bayes* dan dependen yakni dengan *Long Short Term Memory* (LSTM). Perancangan model sentimen analisis kali ini menggunakan studi kasus pada *brand* Indomie. Kedua model dilatih dan dites dengan menggunakan *datasets* yang serupa, dan hasil klasifikasi kedua model tersebut dibandingkan.

Berdasarkan hasil eksperimen yang telah dilakukan, model *Long Short Term Memory* (LSTM) berhasil mengungguli kinerja *Naïve Bayes*. LSTM berhasil meraih akurasi sebesar: 77.92%, sementara *Naïve Bayes* hanya mencapai: 66.31%. Unggulnya model LSTM dalam menganalisa sentimen *tweets* menandakan bahwa kata-kata di dalam *tweets* bersifat dependen dan sangat mempengaruhi makna suatu *tweets*.

Kata kunci: *naïve bayes, long short term memory, sentimen analisis, study case brand indomie, tweets*

SENTIMENT ANALYSIS COMPARISON TOWARDS INDOMIE BRAND USING NAÏVE BAYES & LONG SHORT TERM MEMORY (LSTM)

ABSTRACT

Oleh: Elisabeth

Sentiments analysis or opinion mining is a field of study that analyzes the opinions, sentiments, evaluation, attitudes, and emotions of its users in written text (Liu B. , 2012). Sentiment analysis is also often discussed in the realm of research (Vinodhini, 2012). Previous research related to the classification of tweets sentiment using Naïve Bayes (Tunggawan, 2016). However, Naïve Bayes is a method that assumes each word frequency independently (Ginosar, 2012) and is more appropriately implemented in datasets with minimal data variation (Pang, 2002). Meanwhile, relationships between tweets tend to be dependent and have wide variations of data with large numbers (Jiang, 2011).

Therefore, further exploration of the previous research is required (Jiang, 2011). This can be done by designing 2 models that assume the data independently with Naïve Bayes and dependently using Long Short Term Memory (LSTM). This time, we will use Indomie brand as our study case. Both models were trained and tested using similar datasets, and the results of the classification of the two models were compared.

Based on the results of the experiments, Long Short Term Memory (LSTM) model successfully outperformed the Naïve Bayes performance. LSTM achieved accuracy of: 77.92%, while Naïve Bayes only reached: 66.31%. The result indicates that the words in tweets are dependent and greatly affect the meaning of tweets.

Keywords: *naïve bayes, long short term memory, sentiment analysis, study case indomie brand, tweets*

UMMN

KATA PENGANTAR

Puji dan syukur dihaturkan kepada Tuhan Yang Maha Esa sehingga skripsi yang berjudul “Perbandingan Sentimen Analisis terhadap *Brand Indomie* Menggunakan *Naïve Bayes & Long Short Term Memory (LSTM)*” dapat selesai tepat pada waktunya. Skripsi ini diajukan kepada Program Strata 1, Program Studi Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Multimedia Nusantara.

Dengan berakhirnya proses penulisan skripsi ini, ungkapan terima kasih disampaikan kepada Universitas Multimedia Nusantara yang telah memberi beasiswa kepada penulis selama menempuh pendidikan di Universitas Multimedia Nusantara sehingga dapat membantu proses penulisan skripsi dalam membayar biaya kuliah hingga selesai.

Selain itu, ungkapan terima kasih juga disampaikan kepada :

1. Bapak Yustinus Eko Soelistio, S.Kom., M.M. selaku dosen tetap di Universitas Multimedia Nusantara serta sebagai dosen pembimbing skripsi yang telah memberikan bimbingan, informasi, dan nasihat dalam penyusunan laporan skripsi,
2. Ibu Ririn Ikana Desanti, S.Kom., M.Kom. selaku Kepala Program Studi Sistem Informasi Universitas Multimedia Nusantara,

3. Elvyna Tunggawan, S.Kom., selaku alumni Program Studi Sistem Informasi Universitas Multimedia Nusantara yang sudah meluangkan waktunya untuk berdiskusi mengenai perancangan model skripsi.

Tidak lupa penulis mengucapkan terima kasih kepada orang tua, keluarga, dan semua rekan-rekan yang telah memberikan semangat dan doa kepada penulis dalam penyelesaian skripsi ini.

Semoga skripsi ini dapat memberikan informasi dan inspirasi yang bermanfaat bagi para pembaca.

Tangerang, 03 Januari 2018



Elisabeth

DAFTAR ISI

PERNYATAAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	ii
ABSTRAK	iii
ABSTRACT.....	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
DAFTAR RUMUS	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah	3
1.4. Tujuan dan Manfaat.....	4
BAB II LANDASAN TEORI	5
2.1. <i>Naïve Bayes</i>	5
2.2. <i>Recurrent Neural Network</i>	8
2.3. <i>K-fold Cross Validation</i>	9
2.4. <i>Receiver Operating Characteristic</i>	10
2.5. <i>Bag-of-words</i>	11
2.6. <i>Add-One Laplace Smoothing</i>	11
2.7. <i>Word Vector</i>	11

2.8. <i>Long Short Term Memory (LSTM)</i>	12
2.9. <i>Word2vec: Skip-gram model</i>	13
2.10 Penelitian Terdahulu.....	15
BAB III METODE PENELITIAN	19
3.1. Objek dan Variabel Penelitian	19
3.2. Metode Penelitian	19
3.2.1. Pengumpulan Data.....	20
3.2.2. Perancangan Model.....	23
3.2.3. Evaluasi Model	32
3.3. Teknik Olah Data.....	32
BAB IV ANALISA DATA DENGAN NAÏVE BAYES	34
4.1. Analisa Data	34
4.2. Analisa Hasil Klasifikasi Data dengan <i>Naïve Bayes</i>	37
BAB V ANALISA DATA DENGAN LSTM	51
5.1 Analisa Hasil Tes pada Konfigurasi LSTM	51
5.2 Analisa Hasil Klasifikasi Data dengan LSTM	55
5.3 Diskusi.....	63
BAB VI KESIMPULAN & SARAN	74
6.1. Kesimpulan.....	74
6.2. Saran.....	76
DAFTAR PUSTAKA	77
LAMPIRAN	81

DAFTAR GAMBAR

Gambar 2.2. Ilustrasi grafik ROC	10
Gambar 2.3 Skema pada model <i>skip-gram</i>	14
Gambar 2.1. Skema looping pada model LSTM.....	16
Gambar 3.1. Deskripsi langkah penelitian	20
Gambar 3.2. Arsitektur model LSTM	28
Gambar 4.1. Frekuensi kata pada <i>tweets</i> positif.....	35
Gambar 4.2. Frekuensi kata pada <i>tweets</i> negatif.....	36
Gambar 4.3. Grafik ROC <i>Naïve Bayes</i> tanpa <i>Add-One Laplace Smoothing</i> dan dengan kapitalisasi kata untuk <i>tweets</i> positif	38
Gambar 4.4. Grafik ROC <i>Naïve Bayes</i> tanpa <i>Add-One Laplace Smoothing</i> dan dengan kapitalisasi kata untuk <i>tweets</i> negatif	39
Gambar 4.5. Grafik ROC <i>Naïve Bayes</i> tanpa <i>Add-One Laplace Smoothing</i> dan tanpa kapitalisasi kata untuk <i>tweets</i> positif.....	40
Gambar 4.6. Grafik ROC <i>Naïve Bayes</i> tanpa <i>Add-One Laplace Smoothing</i> dan tanpa kapitalisasi kata untuk <i>tweets</i> negatif.....	41
Gambar 4.7. Grafik ROC <i>Naïve Bayes</i> dengan <i>Add-One Laplace Smoothing</i> dan dengan kapitalisasi kata untuk <i>tweets</i> positif	42
Gambar 4.8. Grafik ROC <i>Naïve Bayes</i> dengan <i>Add-One Laplace Smoothing</i> dan dengan kapitalisasi kata untuk <i>tweets</i> negatif	43
Gambar 4.9. Grafik ROC <i>Naïve Bayes</i> dengan <i>Add-One Laplace Smoothing</i> dan tanpa kapitalisasi kata untuk <i>tweets</i> positif	44

Gambar 4.10. Grafik ROC <i>Naïve Bayes</i> tanpa <i>Add-One Laplace Smoothing</i> dan tanpa kapitalisasi kata untuk <i>tweets</i> negatif	45
Gambar 5.1. Persebaran <i>Weight</i> pada Salah Satu Sesi <i>Training LSTM</i>	58
Gambar 5.2. Nilai <i>Loss</i> yang Stagnan pada Model LSTM	65
Gambar 5.3. Distribusi <i>weight</i> pada eksperimen pertama dengan 10 <i>layers</i> (13,715 <i>tweets</i>)	66
Gambar 5.4. <i>Loss</i> pada eksperimen pertama dengan 10 <i>layers</i> (13,715 <i>tweets</i>)...	67
Gambar 5.5. Distribusi <i>weight</i> pada eksperimen kedua dengan 10 <i>layers</i> (3,412 <i>tweets</i>)	68
Gambar 5.6. <i>Loss</i> pada eksperimen kedua dengan 10 <i>layers</i> (3,412 <i>tweets</i>).....	69
Gambar 5.7. Distribusi <i>weight</i> pada eksperimen ketiga dengan 10 <i>layers</i> (6,824 <i>tweets</i>)	70
Gambar 5.8. <i>Loss</i> pada eksperimen ketiga dengan 10 <i>layers</i> (6,824 <i>tweets</i>).....	71
Gambar 5.9. Distribusi <i>weight</i> pada eksperimen keempat dengan 1 <i>layer</i> (6,824 <i>tweets</i>)	72
Gambar 5.10. <i>Loss</i> pada eksperimen keempat dengan 10 <i>layers</i> (6,824 <i>tweets</i>)..	72



DAFTAR TABEL

Tabel 4.1. Hasil tingkat akurasi <i>Naïve Bayes</i> tanpa fitur <i>Add-One Laplace Smoothing</i> dengan <i>datasets</i> proporsional.....	46
Tabel 4.2. Hasil tingkat akurasi <i>Naïve Bayes</i> dengan fitur <i>Add-One Laplace Smoothing</i> dengan <i>datasets</i> proporsional.....	47
Tabel 5.1. Hasil Tes Konfigurasi Model LSTM	52
Tabel 5.2. Hasil Akurasi LSTM dengan <i>Datasets</i> Tidak Proporsional.....	57
Tabel 5.3. Hasil Akurasi LSTM <i>Datasets</i> Proporsional dengan 10 <i>Layers</i> (3,412 <i>Tweets</i>)	59

UMMN

DAFTAR RUMUS

Rumus 2.1. Perhitungan model <i>Multinomial</i>	5
Rumus 2.2. Perhitungan <i>Positive Prediction Accuracy</i>	6
Rumus 2.3. Perhitungan <i>Negative Prediction Accuracy</i>	7
Rumus 2.4. Perhitungan <i>Total Accuracy</i>	7
Rumus 2.5. Perhitungan <i>Certainty</i>	8
Rumus 2.6. Perhitungan <i>hidden layer & output LSTM</i>	9
Rumus 2.7 Perhitungan pada sel memori LSTM	12

UMMN