

Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM)

Jovita Nurvania¹, Jondri², Kemas Muslim Lhaksamana³

^{1,2,3} Universitas Telkom, Bandung

¹jnurvania@students.telkomuniversity.ac.id, ²jondri@telkomuniversity.ac.id,

³kemasmuslim@telkomuniversity.ac.id

Abstrak

Berkembangnya teknologi khususnya media sosial dan internet menjadikan keduanya sebagai kebutuhan pokok bagi masyarakat, karena dapat membantu masyarakat dalam mendapatkan informasi dan menyebarkan informasi. Salah satu informasi yang diperlukan oleh masyarakat adalah informasi mengenai sebuah tempat wisata. Media yang berkontribusi besar dalam penyebaran informasi ini adalah situs web. Dalam penyebarannya, informasi dapat dibagi menjadi dua jenis yaitu informasi yang negatif, maupun positif. Analisis sentimen digunakan dalam pengolahan paragraf yang berisi kalimat menggunakan bahasa sehari-hari manusia menjadi bahasa komputer. Penelitian ini bertujuan untuk mengklasifikasikan ulasan pengunjung tentang pengaruh COVID-19 terhadap tempat wisata di Bali dari Tripadvisor menggunakan metode Long Short-Term Memory (LSTM). Sebelum diproses dengan LSTM, setiap teks pada ulasan akan divektorisasi dengan word2vec. Hasil pengujian pada model yang dibangun didapatkan nilai akurasi sebesar 71,67%,

Kata kunci : wisata, situs web, ulasan, analisis sentimen, LSTM, word2vec, COVID-19 di Bali, Tripadvisor

Abstract

The development of technology in particular social media and internet make both of those as people basic needs, because they can help people to reach out and spread information. One of the important information that needed by people is information about tourist attraction. The biggest media that contribute in spreading information is website. In the spreading process, information divided to two kind, they are negative information and positive information. Sentiment Analysis is used to processing a paragraph that consist of human daily language becomes computer language. This research is intend to classifying people's review about COVID-19 impact toward tourist attraction at Bali from Tripadvisor using Long Short-Term Memory (LSTM) method. Before it processed with LSTM, every review text would be vectorized with Word2Vec. The result of model testing obtained accuration score as much as 71,67%.

Keywords: tourist attraction, website, review, sentiment analysis, LSTM, word2vec, COVID-19 in Bali, Tripadvisor

1. Pendahuluan

Latar Belakang

Seiring dengan berkembangnya teknologi terutama internet, internet kian menjadi bagian terpenting bagi masyarakat dunia. Menurut [1] yang menyebabkan internet berkembang salah satunya adalah world wide web (www atau website), dimana dengan hanya mengklik link yang ada langsung terhubung dengan jutaan database yang berisi teks, suara, video, animasi dan lainnya. Dapat diartikan bahwa masyarakat dunia sangat membutuhkan website sebagai sarana mendapatkan informasi, berbagi informasi maupun bertukar informasi ke sesama pengguna website. Dengan banyaknya pengguna yang saling bertukar informasi, menjadikan informasi sebuah hal penting dan sangat riskan jika salah dalam penerimaannya. Karena pada website pengguna bebas mengutarakan pendapatnya baik itu positif maupun negatif, di website pula pengguna dapat dengan bebas mendapatkan informasi lain dari pengguna lainnya.

Salah satu informasi yang dicari oleh para pengguna internet adalah informasi mengenai tempat wisata [1]. Mengetahui keadaan suatu tempat wisata sebelum mendatanginya adalah hal yang sangat penting agar tidak terjadinya kekecewaan atau penyesalan setelahnya. Oleh karena itu, diperlukannya pencarian informasi mengenai suatu tempat wisata yang akan dikunjungi [1]. Salah satu *website* yang menyediakan berbagai informasi mengenai destinasi wisata Tripadvisor. Menurut [7] Tripadvisor sangat populer di kalangan wisatawan dan pengelola tempat wisata, yang menjadikan Tripadvisor menjadi tempat menulis berbagai ulasan dari

Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode *Long Short-Term Memory* (LSTM)

Juni 2021

destinasi wisata di seluruh dunia. Oleh karena itu pada [2] dilakukan sentiment analysis pada ulasan sebuah travel karena sentiment analysis dapat mengekstraksi sebuah teks dari sebuah tempat tertentu dan kemudian dapat diklasifikasikan ke positif maupun negatif tergantung isinya. Pada [2] penelitian juga merepresentasikan dokumen menjadi vektor yang kemudian diolah dengan metode Naive Bayes, Support Vector Machine (SVM) dan N-gram.

Penelitian kali ini menerjemahkan pendapat para wisatawan yang menulis ulasan pada situs web, diperlukannya algoritma untuk memproses informasi tersebut dan memetakannya menjadi sebuah vektor menggunakan metode word2vec dan klasifikasi menggunakan LSTM. Karena dengan menggunakan vektor sebuah kalimat dapat diproses menggunakan algoritma pada natural language processing (NLP) [3]. NLP digunakan untuk analisis sentimen agar mengetahui apakah sebuah dokumen teks termasuk sentimen positif atau sentimen negatif. Maka dari itu, metode LSTM digunakan pada penelitian ini untuk analisis sentimen ulasan yang ada pada forum website TripAdvisor.

Topik dan Batasannya

Topik yang dibahas pada penelitian kali ini adalah berita keadaan di Bali di tengah pandemic COVID-19 pada forum Tripadvisor. Dataset yang digunakan berasal dari website Tripadvisor dengan jumlah dataset sebanyak 600 dan pembagian sebanyak 80% untuk data latih dan 20% untuk data uji yang diunggah dari bulan Maret 2020 hingga Januari 2021. Metode yang digunakan adalah *Long Short-Term Memory* (LSTM) dengan sebelumnya data melalui *preprocessing* dan ekstraksi dataset menggunakan Word2Vec. *Preprocessing* yang digunakan adalah *data cleansing*, *stopword removal*, *lower casing*, *tokenization* dan *stemming*.

Tujuan

Tujuan dari penelitian ini adalah membuat sistem klasifikasi data ulasan pada *website* TripAdvisor menggunakan model *Long Short-Term Memory* (LSTM). Tujuan lain dari penelitian ini adalah untuk mengetahui bagaimana model LSTM dapat mengklasifikasikan sebuah dokumen ulasan tempat wisata, dan menganalisis performansi LSTM yang didapatkan dari hasil klasifikasi dokumen ulasan tempat wisata. Serta mengetahui pengaruh *preprocessing data* dan *undersampling data* pada performansi LSTM.

Organisasi Tulisan

Bagian selanjutnya adalah bab 2 yang akan membahas studi terkait tentang penelitian yang dilakukan. Kemudian bab 3 yang akan membahas sistem yang dibangun pada penelitian ini. Selanjutnya hasil dan analisis penelitian ada pada bab 4, dan bab 5 membahas kesimpulan. Bagian akhir berisi referensi atau daftar pustaka mengenai pengerjaan penelitian ini.

2. Studi Terkait

Sebelum dilakukannya pengklasifikasian menggunakan LSTM, dokumen teks harus diubah menjadi bentuk vektor. Seperti penelitian yang dilakukan pada jurnal [12] dinyatakan bahwa penggunaan word2vec untuk mengubah teks menjadi vektor dapat menghasilkan akurasi yang tinggi. Word2vec membantu komputer untuk mengidentifikasi kombinasi kata yang ada. Vektor juga dibutuhkan karena dapat merepresentasikan setiap kata menjadi satu vektor agar lebih mudah dibaca oleh komputer [14]. Dibuktikan pada [12] dengan menggunakan vektor akurasi pada sebuah klasifikasi dapat mencapai 72% dibantu dengan Support Vector Classifier.

Selain itu, dapat juga dilihat pada penelitian yang dilakukan oleh Qiang Ye, Ziqiong Zhang dan Rob Law pada [2] dimana penelitian tersebut melakukan sentiment analysis pada sebuah travel blog untuk tujuan travel tertentu. Penelitian ini menggunakan tiga metode yaitu Naive Bayes, SVM dan N-gram, dan sebelum diklasifikasikan dokumen direpresentasikan ke dalam bentuk vektor. Untuk mengevaluasi hasil pengerjaannya pun menggunakan K-fold cross validation dengan menggunakan 600 ulasan positif dan 591 ulasan negatif. Hasil akurasi dari ketiga metode tersebut sangat tinggi yaitu lebih dari 80%. Tetapi saat dicoba beberapa data saja dapat dilihat bahwa semakin banyak data latih yang dipakai nilai akurasi akan semakin tinggi.

Penelitian lainnya pernah dilakukan oleh Fenna Miedema yang menggunakan data ulasan film di situs IMDB dengan menggunakan 50,000 dataset dengan 25,000 untuk data latih dan 25,000 data validasi dengan masing-masing jumlah kelas positif sebanyak 50% dan negatif sebanyak 50%. Penelitian tersebut menggunakan LSTM dan memiliki nilai akurasi yaitu 86,75% [4]. Penelitian menggunakan text mining dan sentiment analysis lainnya juga dilakukan oleh Haruna Isah, Paul Trundle dan Daniel Neagu yang menggunakan media sosial sebagai sumber datasetnya. Dataset diambil dari twitter dengan jumlah total 11,431 tweets dengan data latih sebanyak 75% dan data uji 25%. Pengujian dilakukan menggunakan metode Naive Bayes dengan akurasi sebanyak 83% [5].

Penelitian lain yang pernah dilakukan adalah penelitian data dari weibo oleh bai xue, chen fu dan zhan shaobin. Penelitian ini mengambil data dari weibo dan mengklasifikasikan dengan word2vec. Dataset diambil dari konten dari weibo yang dibagi menjadi tiga kelas. Dari keseluruhan dataset 70% dijadikan data latih dengan

Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode *Long Short-Term Memory* (LSTM)

Juni 2021

word2vec, dan 30% sisanya dijadikan data validasi keefektifan dan akurasi dari model yang telah dibangun. Hasil dari penelitian ini mendapatkan nilai akurasi sebanyak 83%, hasil model tersebut dapat lebih tinggi akurasinya dengan menggunakan *high sentiment value* [8].

Text Mining

Text mining atau penambangan teks adalah sebuah proses untuk mendapatkan informasi dengan melakukan penggalian pola-pola dari data agar data menjadi informasi yang lebih berharga [5]. Teknik ini dilakukan pada sebuah database yang berbasis teks [6] baik berupa dokumen word, pdf maupun kutipan teks. Penambangan teks berguna untuk mengekstraksi sebuah informasi yang awalnya berupa data atau dokumen yang tidak berpola seperti data yang ada pada sosial media. Text mining dapat dikatakan menggunakan teknik mirip dengan natural language processing (NLP) dalam mendapatkan informasi dan mengekstrak informasi [6].

Feature Extraction

Pada tahapan ini, data diubah menjadi bentuk vektor. Feature Extraction memiliki dua fase yang pertama adalah ekstraksi primer yang mengubah data menjadi format standar kemudian baru diubah menjadi bentuk vektor [10]. Seperti pada penelitian ini, ulasan yang ada pada TripAdvisor akan diubah menjadi bentuk vektor. Feature extraction yang digunakan untuk mengubah data ke bentuk vektor yaitu dengan metode Word2Vec. Pada website [9] dijelaskan bahwa word2vec adalah sebuah metode untuk merepresentasikan setiap kata pada dokumen ke dalam bentuk vektor dengan N dimensi. Word2vec membutuhkan masukan berupa kumpulan teks dan mengubah kata-kata di dalamnya menjadi vektor sebagai keluarannya [8]. Hal ini bertujuan agar komputer lebih mudah menjelaskan makna dari kata tersebut. Hasil dari vektor tersebut dapat diaplikasikan pada banyak metode natural language processing dan machine learning. Pada [14] dinyatakan bahwa word2vec mempunyai atribut lebih yang tidak ada pada klasifikasi teks yang lain, hal ini membuat word2vec sangat bermanfaat.

Recurrent Neural Network (RNN)

Metode Recurrent Neural Network (RNN) pertama kali dijalankan oleh Mikolov karena dikenal dapat memproses data teks yang berurutan [16]. RNN adalah jaringan yang berisi perulangan, dimana informasi bisa saja terduplikasi pada satu jaringan setiap kali informasi ini melewati jaringan [11]. RNN memiliki tiga layer yaitu input layer, hidden layer dan output layer [16]. RNN merupakan model dari saraf jaringan yang prosesnya tidak bergantung pada ukuran window, RNN menyimpan semua masukannya pada hidden layer yang berisi kata masukkan dan kata selanjutnya [10]. Oleh karena itu, menurut jurnal [15] RNN sangat menjanjikan dalam menerjemahkan bahasa mesin. Contohnya adalah jika terdapat empat kata maka akan terdapat empat layer saraf jaringan juga karena setiap satu layernya merepresentasikan satu kata.

Pada [15] dikatakan bahwa sebenarnya kita dapat menggunakan Backpropagation Through Time (BPTT) untuk mengecek hasil dari RNN, tetapi akan sulit karena RNN tidak dapat menyimpan sebuah data yang panjang. Menurut [11] juga RNN tidak dapat menghubungkan informasi jika adanya jarak pada sebuah informasi atau rangkaian kata. Maka dari itu, karena RNN tidak dapat menyelesaikan masalah tersebut diciptakan metode baru bernama Long Short-Term Memory (LSTM) yang memiliki keunggulan dalam menangani informasi yang panjang dengan efisien [4].

Long Short-Term Memory (LSTM)

Metode LSTM pertama kali dibuktikan oleh Hochreiter and Schmidhuber pada tahun 1997 bahwa metode ini dapat mengolah data yang panjang [15]. LSTM adalah versi terbaru yang lebih efisien dan memiliki keuntungan yang lebih dibandingkan dengan RNN [13]. Pada LSTM meskipun terdapat jarak antar teks analisis masih tetap bisa dilakukan. Hal ini terjadi karena pada LSTM terdapat memori yang menyimpan informasi sebelumnya, informasi dapat disimpan untuk waktu yang lama [10]. Jaringan LSTM memiliki ukuran yang besar, pada jaringan ini terdapat LSTM cell yang menggantikan posisi hidden layer pada RNN, gunanya untuk menyimpan teks sebelumnya [16]. Pada jaringan LSTM terdapat *forget gate*, *input gate*, *output gate* dan *memory cell* yang akan menghitung nilai keluaran sebagai hidden layer untuk jaringan selanjutnya [4]. Berikut adalah rumus dari setiap *gate* di LSTM :

Analisis Sentimen Pada Ulasan di TripAdvisor
Menggunakan Metode *Long Short-Term Memory*
(*LSTM*)

Juni 2021

Confusion Matrix

Confusion Matrix adalah sebuah tabel atau matriks yang berisikan empat nilai yang merupakan pengukuran performa dari masalah klasifikasi yang telah dilakukan. Ada empat nilai atau point yang ada pada confusion matrix yaitu True Positive, True Negative, False Positive dan False Negative. Gambaran confusion matrix dapat dilihat pada tabel berikut [11].

Table 1. Gambaran confusion matrix

		True Values	
		Positive	Negative
Predictions	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Keterangan :

- True positive (TP) : Prediksi yang bernilai positive dan benar sesuai target.
- True negative (TN) : Prediksi yang bernilai negative dan benar sesuai target.
- False positive (FP) : Prediksi yang bernilai positive dan salah tidak sesuai target.
- False negative (FN) : Prediksi yang bernilai negative dan salah tidak sesuai target.

Setelah diketahui confusion matrixnya, dapat diketahui juga nilai *accuracy*, *precision*, *recall* dan *f1-score*. Berikut adalah penjelasan dan rumus untuk mengetahuinya :

- *Accuracy*

Accuracy adalah perhitungan seberapa tepat klasifikasi yang telah dibangun, sesuai dengan target yang ada [11].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

- *Precision*

Precision adalah perhitungan keakuratan antara data target dengan hasil prediksi dari model [11].

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$$

- *Recall*

Recall adalah perhitungan yang menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi [11].

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%$$

- *F1-Score*

F1-score adalah perhitungan yang menggambarkan perbandingan antara *precision* dan *recall*. Jika nilai FN dan FP tidak mendekati sebaiknya digunakan *f1-score* dibandingkan nilai akurasi [11].

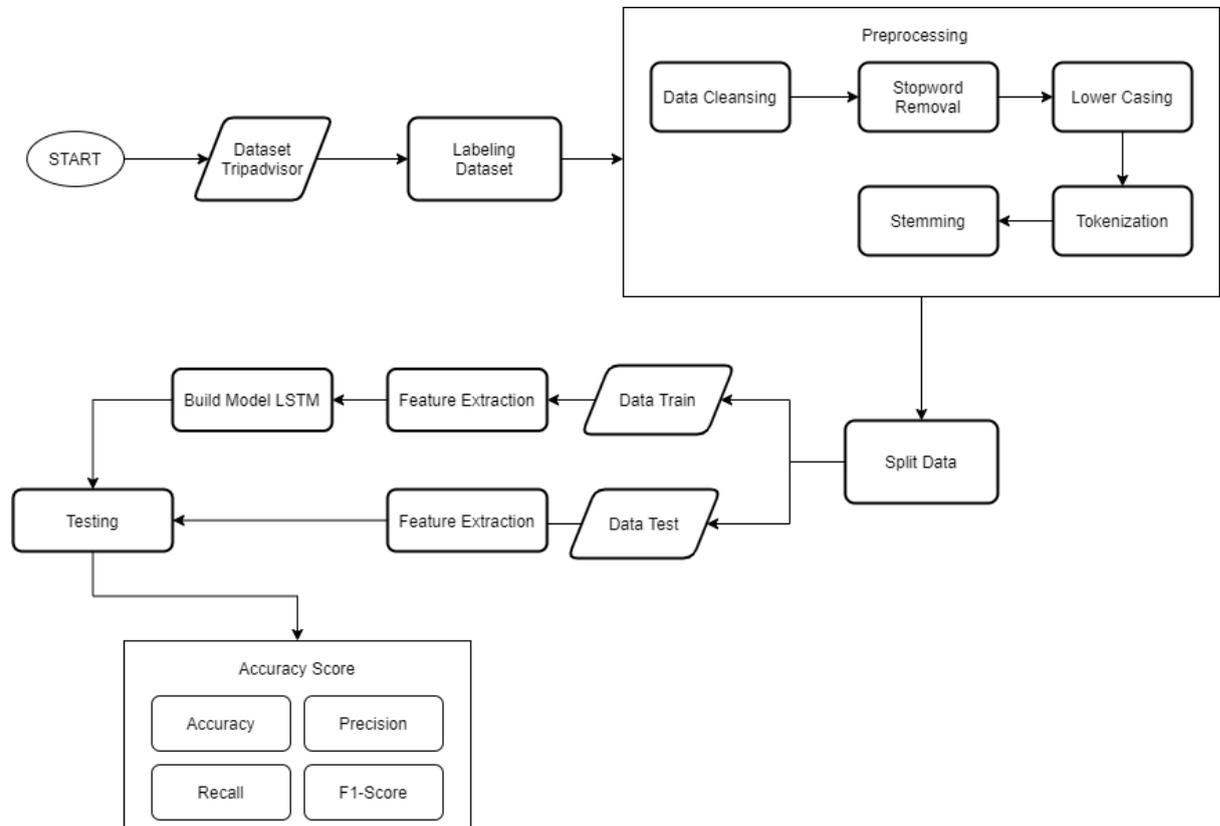
$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%$$

Analisis Sentimen Pada Ulasan di TripAdvisor
Menggunakan Metode *Long Short-Term Memory*
(*LSTM*)

Juni 2021

3. Sistem yang Dibangun

Rancangan sistem yang dibangun pada tugas akhir ini adalah sistem yang dapat mengklasifikasikan ulasan yang ada pada Tripadvisor mengenai perkembangan virus COVID-19 di Bali ke dalam kelas negatif maupun positif. Adapun beberapa tahapan yang harus dilakukan untuk pengklasifikasian tersebut, tahapan-tahapannya adalah pengambilan data, *preprocessing*, *feature extraction*, *classification*, dan *evaluation*. Tahapan-tahapan tersebut bila dibentuk *flowchart* sesuai dengan urutannya adalah sebagai berikut.



Gambar 1. Rancangan umum sistem

Dataset

Pengambilan dokumen atau *dataset* yang digunakan dalam penelitian Tugas Akhir ini berasal dari website Tripadvisor yang berisi pertanyaan dari turis asing yang akan mengunjungi pulau Bali tetapi terhalang oleh pandemic COVID-19 dan jawaban atau ulasan baik dari masyarakat sekitar maupun masyarakat yang mengetahui informasi tentang pulau Bali. Data tersebut akan digunakan menjadi data latih maupun data uji, jumlah *dataset* keseluruhan berjumlah 600 dimana untuk data latih menggunakan 80% dari dataset dan data uji menggunakan 20% dari dataset. Kemudian, untuk pembagian kelas pada keseluruhan data masing-masing data yang berkelas positif berjumlah 411 dan kelas negatif berjumlah 189. Dengan jumlah masing-masing kelas yang telah disebutkan, maka dataset dianggap tidak seimbang. Oleh karena itu, pada dataset dilakukan *undersampling* atau penyeimbangan dataset dengan mengurangi ukuran kelas yang lebih banyak.

Tabel 2. Contoh dataset yang akan diolah dan sudah diberi label

Label	Kalimat
Negatif	We are also travelling to Bali on the same day via Bangkok and really worried about the situation.
Positif	I can only hope for the best and Im happy at least they are doing something.

Analisis Sentimen Pada Ulasan di TripAdvisor
Menggunakan Metode *Long Short-Term Memory*
(*LSTM*)

Juni 2021

Labeling Dataset

Pelabelan dataset dilakukan secara manual maupun bantuan dari website <http://sentistrength.wlv.ac.uk/>. Karena hanya terdapat dua label, maka data yang termasuk ke dalam label netral dimasukkan ke label positif. Data yang termasuk ke dalam kelas positif adalah data yang berisi ulasan keadaan Bali maupun perbandingan dengan daerah lain yang menyatakan bahwa pengaruh dari pandemi COVID-19 ini baik baik saja tidak menyebabkan hal buruk. Sedangkan, untuk data yang termasuk ke dalam kelas negatif adalah data yang berisi kekhawatiran masyarakat terhadap kondisi COVID-19 di banyak wilayah terutama di Bali serta pengaruh buruk yang disebabkan oleh COVID-19.

Preprocessing

1. Data Cleansing

Pada tahap awal ini, data teks yang sudah ada dibersihkan dengan menghilangkan simbol, angka, tanda baca, spasi yang berlebihan maupun karakter yang tidak ada di alfabet. Dan juga mengubah kata singkatan menjadi kata aslinya.

Tabel 3. Preprocessing Data Cleansing

Kalimat	Data cleansing
Hello Vishal, If you believe the latest statistics being 3 cases in India, 35 cases in Thailand and still none in Bali. You are safer in Bali than anywhere else. I'm in Bali atm and life is as normal here.	Hello Vishal If you believe the latest statistics being cases in India cases in Thailand and still none in Bali You are safer in Bali than anywhere else I am in Bali atm and life is as normal here

2. Stopwords Removal

Tahap ini berfungsi untuk menghapus kata yang tidak penting dan tidak memiliki makna untuk proses selanjutnya, atau menghapus kata yang sering berulang [17].

Tabel 4. Preprocessing Stop Words Removal

Kalimat	Stop word removal
Hello Vishal If you believe the latest statistics being cases in India cases in Thailand and still none in Bali You are safer in Bali than anywhere else I am in Bali atm and life is as normal here	Hello Vishal If believe the latest statistics being cases India cases Thailand still none Bali safer Bali anywhere else Bali atm life normal

3. Lower Casing

Pada tahap ini semua huruf yang ada disetarakan bentuknya yaitu menjadi huruf kecil atau huruf besar semua, tujuannya untuk mempermudah proses setelahnya.

Tabel 5. Preprocessing Lower Casing

Kalimat	Lower casing
Hello Vishal If believe the latest statistics being cases India cases Thailand still none Bali safer Bali anywhere else Bali atm life normal	hello vishal if believe the latest statistics being cases india cases thailand still none bali safer bali anywhere else bali atm life normal

4. Tokenization

Tahap ini berperan memisahkan teks dokumen menjadi serangkaian token atau kata [17].

Tabel 6. Preprocessing Tokenization

Kalimat	Tokenization
hello vishal if believe the latest statistics being cases india cases thailand still none bali safer bali anywhere else bali atm life normal	['hello', 'vishal', 'if', 'believe', 'the', 'latest', 'statistics', 'being', 'cases', 'india', 'cases', 'thailand', 'still', 'none', 'bali', 'you', 'safer', 'bali', 'anywhere', 'else', 'bali', 'atm', 'life', 'normal']

5. Stemming

Pada tahap ini dilakukan penghapusan imbuhan baik di awal (prefixes), di akhir (suffixes), kata sisipan (infixes) maupun imbuhan di awal kata dan di akhir (confixes) pada setiap kata sehingga kata tersebut menjadi kata dasar [17].

Tabel 7. Preprocessing Stemming

Kalimat	Stemming
['hello', 'vishal', 'if', 'believe', 'the', 'latest', 'statistics', 'being', 'cases', 'india', 'cases', 'thailand', 'still', 'none', 'bali', 'safer', 'bali', 'anywhere', 'else', 'bali', 'atm', 'life', 'normal']	['hello', 'vishal', 'if', 'believ', 'latest', 'statist', 'case', 'india', 'case', 'thailand', 'still', 'none', 'bali', 'safer', 'bali', 'anywher', 'els', 'bali', 'atm', 'life', 'normal']

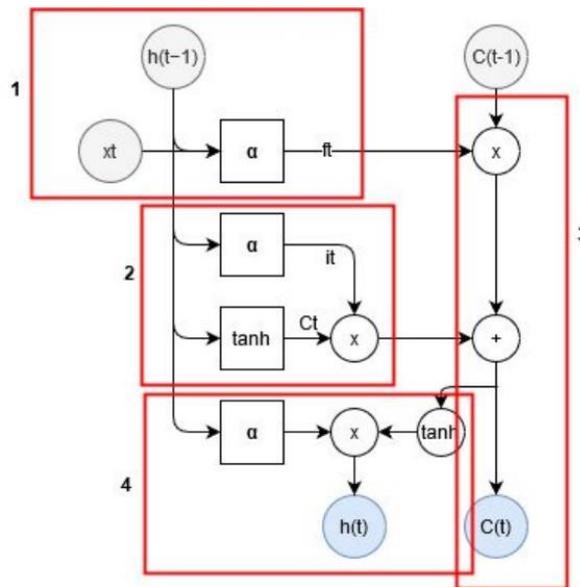
Feature Extraction

Dalam rancangan sistem yang telah dibuat, setelah dilakukannya proses preprocessing maka tahapan selanjutnya adalah feature extraction. Pada tahap ini sistem akan mengubah bentuk sebuah kata menjadi vektor menggunakan model word2vec, yang nantinya akan mempermudah dalam melewati tahapan klasifikasi menggunakan Long Short-Term Memory (LSTM). Model word2vec yang digunakan adalah Skip-Gram Negative Sampling (SGNS). Menurut [18] metode Skip-Gram memiliki kelebihan yaitu dapat mendeteksi kata-kata dasar yang jarang muncul, sehingga penggunaan metode ini lebih tepat dibandingkan dengan CBOW. Dan pada [21] dibuktikan bahwa model Skip-Gram memiliki performansi yang paling baik. Berdasarkan penelitian didapatkan bahwa model *Negative Sampling* mengalahkan *Hierarchical Softmax*.

Setiap kata yang ada dibuat model vektornya dari word2vec. Kata tersebut dengan menggunakan one-hot coding merepresentasikan kata sebagai vektor $|V|$ dengan begitu semua nilai dari kata yang ada di konteksnya akan 0 atau random kecuali kata yang cocok dengan input [19]. Pada penelitian kali ini, yang digunakan adalah sekumpulan data yang berisi *corpus* kata yang berasal dari Wikipedia dengan dimensi 300[22]. Karena pada [22] dikatakan bahwa vektor dengan dimensi 300 meningkatkan performansi model dibandingkan dengan vektor dengan dimensi 100.

Klasifikasi Long Short-Time Memory (LSTM)

Setelah tahap pengekstrasian fitur selesai maka akan dilanjutkan ke tahap klasifikasi, pada penelitian kali ini menggunakan metode Long Short-Term Memory (LSTM) yang telah didefinisikan pada bab sebelumnya. LSTM memproses data yang terstruktur tersebut menggunakan gerbang vektor yang setiap lapisannya mengontrol setiap informasi yang lewat [3]. Pada setiap *time step* di LSTM ada sebuah vektor yang nantinya akan melalui unit LSTM yang terdiri dari *forget gate*, *input gate*, *output gate* dan *memory cell*. Proses ini digunakan untuk menghitung nilai keluaran dari nilai *hidden layer* [4].



Gambar 2. Struktur Jaringan LSTM

Dapat dilihat pada gambar2 di atas pada proses nomor 1, vektor yang masuk sebagai *input* (x_t) akan *diconcatenate* atau digabung dengan vektor dari *output* sebelumnya atau $h(t-1)$. Kemudian gabungan dari kedua vektor tersebut melalui aktivasi sigmoid yang membuat *output* dari *forget gate* merupakan nilai antara 0 hingga 1. Pada proses nomor 2, ada proses yang menciptakan nilai baru dari inputan saat ini (x_t). Pertama-tama vektor input dan vektor output sebelumnya *diconcatenate* dan pada gambar terdapat dua aktivasi, yang pertama diaktivasi dengan sigmoid yg kedua diaktivasi dengan tanh, kemudian kedua hasil aktivasi tersebut dikalikan.

Pada proses nomor 3, ada proses pengkalian nilai *cell state* sebelumnya dengan *forget gate* yang kemudian hasil perkaliannya akan ditambah dengan nilai dari proses nomor 2. Proses nomor 4 adalah proses untuk nilai pada *output gate* yang berasal dari gabungan dari x_t dan $h(t-1)$ yang diaktivasi dengan fungsi sigmoid. Kemudian untuk diteruskan ke lajur *output* sebagai nilai $h(t)$ dikalikan dengan memori yang telah diaktivasi dengan tanh. Nilai tersebut digunakan untuk jaringan lstm selanjutnya, kemudian memori yang tidak diaktivasi dengan tanh diinputkan ke memory lstm selanjutnya.

Testing

Tahap terakhir sebelum dievaluasi adalah pengujian model yang telah dibangun. Pengujian dilakukan dengan memroses data uji yang telah ada untuk mengetahui hasil prediksi label sesuai dengan target yang sudah ada untuk kemudian dihitung nilai ketepatannya.

Evaluasi

Evaluasi yang dilakukan pada penelitian ini adalah dengan membandingkan label prediksi dengan label target atau data aslinya untuk mengetahui confusion matrixnya. Kemudian dihitung seberapa banyak data yang sama untuk diketahui nilai akurasi, precision, recall dan f1-score.

Confusion matrix :

Tabel 8. Hasil confusion matrix

	True Values	
Predictions	81	30
	4	5

Analisis Sentimen Pada Ulasan di TripAdvisor
Menggunakan Metode *Long Short-Term Memory*
(LSTM)

Juni 2021

Dari confusion matrix di atas maka dapat dihitung nilai :

- Akurasi : $\frac{TP + TN}{TP + FP + FN + TN} \times 100\% = 71.67\%$
- Precision : $\frac{TP}{TP + FP} \times 100\% = 72.97\%$
- Recall : $\frac{TP}{TP + FN} \times 100\% = 95,3\%$
- F1-Score : $2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% = 82,7\%$

4. Evaluasi

Hasil Pengujian

Penelitian ini menggunakan *dataset review* dari Tripadvisor sebanyak 600 data dengan pembagian data 80% untuk data latih dan 20% untuk data uji. Setiap data memiliki dua kelas yaitu positif dan negatif. Skenario yang digunakan pada penelitian ini ada sepuluh skenario pengujian.

Tabel 9. Skenario Pengujian

Skenario	Preprocessing	Undersampling	epochs
I	X	X	5
II	X	X	100
III	X	O	5
IV	X	O	100
V	O	X	5
VI	O	X	100
VII	O	O	5
VIII	O	O	100

Dari adanya sepuluh skenario tersebut, berikut adalah hasil perhitungan nilai *accuracy*, *precision*, *recall* dan *f1-score*.

Tabel 10. Hasil Pengujian

Skenario	Hasil			
I	55.83%	70.51%	64.7%	67.5%
II	60.83%	73.17%	70.6%	71.9%
III	60%	72.29%	70.6%	71.4%
IV	61.67%	73.49%	71.8%	72.6%
V	65%	75.29%	75.3%	75.3%
VI	64.17%	75%	74.1%	74.6%
VII	71.67%	72.97%	95.3%	82.7%
VIII	65.83%	75.58%	76.5%	76%

Analisis Hasil Pengujian

Dapat dilihat pada tabel 8 hasil pengujian bahwa skenario yang memiliki nilai akurasi tertinggi adalah skenario VII yaitu dengan akurasi 71.67%, dimana skenario I mempunyai komposisi yaitu data latih melewati *preprocessing*, dilakukan *undersampling* pada data, selain itu juga dilakukan fitting model dengan data latih dan

Analisis Sentimen Pada Ulasan di TripAdvisor
Menggunakan Metode *Long Short-Term Memory*
(*LSTM*)

Juni 2021

label sebanyak 5 kali. Selain nilai akurasi, skenario VII juga memiliki nilai F1-Score yang paling tinggi, hal ini berarti data yang dipakai sudah seimbang dibandingkan dengan yang lain.

Dapat dilihat skenario lainnya yang memiliki atribut sama kecuali pada epochs seperti skenario I dan II, skenario III dan IV, dan skenario V dan VI, serta skenario VII dan VIII. Masing-masing pasangan skenario tersebut memiliki nilai akurasi yang berbeda, dimana skenario dengan epochs lebih besar tidak selalu lebih tinggi nilainya. Hal ini berarti jumlah epochs yang dilakukan tidak mempengaruhi nilai akurasi, precision, recall dan f1-score.

Selanjutnya, untuk data yang tidak melalui *preprocessing* seperti pada skenario I, II, III dan IV. Untuk skenario I dan II selain data tidak melewati *preprocessing*, data juga tidak di *undersampling*, hal ini menyebabkan nilai akurasi yang rendah untuk pengujian kali ini. Untuk skenario III dan IV data hanya tidak melalui *preprocessing* saja dan tetap di *undersampling*, hasilnya sedikit lebih baik dibandingkan dengan skenario I dan II dilihat dari nilai akurasi. Sedangkan untuk skenario V dan VI dimana data hanya melewati proses *preprocessing* tetapi tidak melewati *undersampling* hasilnya lebih baik lagi dari yang sebelumnya. Oleh karena itu, pada perbandingan ketiga pasangan skenario tersebut dapat dilihat bahwa hal terpenting pada pemrosesan data adalah *preprocessing*nya, karna walaupun dengan adanya penyeimbangan data tanpa *preprocessing* hasilnya akan tetap rendah.

Analisis Hasil Klasifikasi

Dari hasil klasifikasi data uji, ada beberapa data yang hasil kelas prediksinya berbeda dengan target kelasnya. Hasil prediksi yang digunakan pada analisis ini adalah prediksi dari skenario yang memiliki nilai akurasi tertinggi, yaitu skenario 7. Berikut adalah penjabarannya;

Tabel 11. Tabel Hasil Klasifikasi

No.	Contoh Kalimat	Target	Prediksi
1.	Glad your 14 days are up Richard. There's pretty dire news coming out of Indonesia these days and sadly fear there's a lot worse to come, very scary.	0	1
2.	The confirmed number of cases in Singapore is increasing daily with I believe 43 cases compared to around 18 cases two or three days ago and it is the country with one of the highest number of cases outside of China. In Indonesia the four confirmed cases are from Indonesians who were in Singapore but escaped to Batam island and currently there is a police chase to find them.	0	1
3.	Every persons who has visited Mainland China in the last 14 days will be refused to enter Indonesia, no matter the nationality. There are tensn have been refused to enter Indonesia. In the big airport, you just walk as usual and thermal scanner will monitore you, maybe you are not realize for it. In small airport, they use manual laser thermal temperature and aim at your forehead.	1	0

Dapat dilihat pada tabel 9 bahwa terdapat tiga data sebagai contoh untuk data yang kelas prediksinya berbeda dengan kelas targetnya. Nilai sentimen pada masing-masing kata dan kalimat yang akan dianalisis berdasarkan nilai sentiment pada *website* *sentistrength*. Perhatikan pada data pertama, kelas target merupakan kelas 0 atau negatif dan memiliki kelas prediksi 1 atau positif. Hal ini terjadi karena pada data terdapat kata *glad* dan *pretty* yang merupakan kata positif serta *news* yang merupakan kata netral dimana pada penelitian ini kelas netral disatukan dengan kelas positif. Hal tersebut membuat hasil prediksi data tersebut masuk ke dalam kelas positif.

Kemudian pada data kedua, kelas target merupakan kelas 0 dan prediksi merupakan kelas 1. Pada data ini terdapat kata *confirmed*, *number*, *case*, *increasing* dan *high* yang sering muncul dan kata-kata tersebut masuk ke dalam kelas netral sehingga hasil prediksinya adalah 1. Pada data terakhir kelas targetnya merupakan kelas 1 dan prediksinya merupakan kelas 0. Hal ini terjadi karena pada data ketiga ini untuk kalimat pertama dan keduanya jika disentimen sendiri-sendiri maka hasil sentimennya adalah negatif. Selain itu juga terdapat kata yang berulang seperti kata *refused* dimana kata ini sentimennya adalah negatif. Hal tersebut dapat menyebabkan hasil prediksi dari data ini adalah sentiment negatif.

5. Kesimpulan

Berdasarkan hasil pengujian dan analisis pengujian yang telah dilakukan dapat disimpulkan bahwa, hal terpenting yang harus diperhatikan adalah tahap *preprocessing* yang dimana dilakukan pembersihan data hingga menjadi data mentah. Jika data tidak melewati pembersihan maka keakuratan data pun bisa rendah. Hal kedua yang harus diperhatikan adalah keseimbangan data. Jadi, jika data tidak seimbang akan sangat mempengaruhi

Analisis Sentimen Pada Ulasan di TripAdvisor
Menggunakan Metode *Long Short-Term Memory*
(LSTM)

Juni 2021

hasil akurasi. Untuk epochs pada fitting data tidak terlalu berpengaruh, selama data yang diolah sudah seimbang dan bersih, dengan epochs kecil atau sebentar pun data akan tetap akurat.

Nilai akurasi dengan menggunakan metode LSTM bisa lebih tinggi lagi jika data yang digunakan jumlahnya lebih banyak. Karena seperti yang sudah dijelaskan bahwa LSTM cocok digunakan untuk mengolah data yang besar. Jadi, saran untuk pengujian selanjutnya adalah menggunakan dataset yang banyak untuk proses pelatihan data seperti jumlah dataset yang digunakan pada jurnal acuan. Selain itu, data yang digunakan harus seimbang dan sudah bersih atau melewati tahap *preprocessing* dengan baik dan benar.

REFERENSI

- [1] Law, R. (2000). Internet in travel and tourism—part I. *Journal of Travel & Tourism Marketing*, 9(3), 65-71.
- [2] Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert systems with applications*, 36(3), 6527-6535.
- [3] K, Dieka Nugraha. 2018. Membuat Model Word2Vec Bahasa Indonesia dari Wikipedia Menggunakan Gensim. [Online] Available at: <https://medium.com/@diekanugraha/membuat-model-word2vec-bahasa-indonesia-dari-wikipedia-menggunakan-gensim-e5745b98714d> [Accessed 31 March 2020].
- [4] Miedema, F. (2018). Sentiment analysis with long short-term memory networks. Research Paper Business Analytics Vrije Universiteit Amsterdam.
- [5] Isah, H., Trundle, P., & Neagu, D. (2014, September). Social media analysis for product safety using text mining and sentiment analysis. In 2014 14th UK workshop on computational intelligence (UKCI) (pp. 1-7). IEEE.
- [6] Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.
- [7] Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*, 32(4), 72-77.
- [8] Xue, B., Fu, C., & Shaobin, Z. (2014, June). A study on sentiment computing and classification of sina weibo with word2vec. In 2014 IEEE International Congress on Big Data (pp. 358-363). IEEE.
- [9] Romadhan, A. 2018. Word2Vec. [Online] Available at: <https://medium.com/@arifromadhan19/word2vec-95c5df46e045> [Accessed 31 March 2020].
- [10] Thomas, M., & Latha, C. A. (2018). Sentimental analysis using recurrent neural network. *International Journal of Engineering & Technology*, 7(2.27), 88-92.
- [11] Bina Nusantara University. (2020). Confusion Matrix. [Online] Available at: <https://socs.binus.ac.id/2020/11/01/confusion-matrix/> [Accessed on April].
- [12] Acosta, J., Lamaute, N., Luo, M., Finkelstein, E., & Andreea, C. (2017). Sentiment analysis of twitter messages using word2vec. *Proceedings of Student-Faculty Research Day, CSIS, Pace University*, 7.
- [13] Klapper-Rybicka, M., Schraudolph, N. N., & Schmidhuber, J. (2001). Unsupervised Learning in Recurrent Neural Networks*. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN 2001)*.
- [14] Lilleberg, J., Zhu, Y., & Zhang, Y. (2015, July). Support vector machines and word2vec for text classification with semantic features. In 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC) (pp. 136-140). IEEE.
- [15] Wang, X., Jiang, W., & Luo, Z. (2016, December). Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In *Proceedings of 19 COLING 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 2428-2437).
- [16] Li, D., & Qian, J. (2016, October). Text sentiment analysis based on long short-term memory. In 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI) (pp. 471-475). IEEE.
- [17] Kadhim, A. I. (2018). An Evaluation of Preprocessing Techniques for Text Classification. *International Journal of Computer Science and Information Security (IJCSIS)*, 16(6).
- [18] Rusli, M., Faisal, M. R., & Budiman, I. (2019). Ekstraksi Fitur Menggunakan Model Word2vec untuk Analisis Sentimen pada Komentar Facebook. *Soliter*, 2, 104-109.
- [19] Xing, C., Wang, D., Zhang, X., & Liu, C. (2014, December). Document classification with distributions of word vectors. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific* (pp. 1-5). IEEE.
- [20] Oinkina. 2015. Understanding LSTM Networks. [Online] Available at: <https://colah.github.io/posts/2015->
- [21] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. dan Dean, J., 2013. Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546.
- [22] Purbalaksone, M.D., 2019. Skip-Gram Negative Sample for Word Embedding in Indonesian Translation Text Classification (Doctoral dissertation, Telkom University).