

Aspect based Sentiment Analysis Aduan Mahasiswa UMSIDA Dimasa Pandemi Menggunakan LSTM

Bayu Anggara Putra¹, Yosi Kristian¹, Esther Irawati Setiawan¹ and Joan Santoso¹

¹ Departemen Teknik Informatika, Fakultas Sains dan Teknologi, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Indonesia

Corresponding author: Bayu Anggara Putra (e-mail: bayu1@mhs.stts.edu).

ABSTRACT The large number of complaint data from Muhammadiyah Sidoarjo University (UMSIDA) students who were affected by the Covid-19 pandemic outbreak, with the implementation of community activity restrictions (PPKM). UMSIDA formed a team named the Umsida Covid-19 Command Center (UCCC), with the aim of implementing a Covid-19 prevention and action program, with the hope that researchers would like to facilitate the delivery of information / student complaints, especially to the UCCC team as consideration in carrying out an activity. decision to deal with the current covid pandemic. Multi aspect sentiment analysis brings something new, to understand user opinions and ratings expressed online. With the aim of classifying subjective texts by labeling polarity, the formation of word vector representation using Word Embedding Global Vector (Glove) is carried out in combination with sentiment analysis training with Long Short Term Memory (LSTM) based classification. Student complaint modeling is done to get vector representation using LSTM. Here, each word of the sentence occupies one LSTM processing step, and the output of the last word is used as the expression of the sentence. The results of the study using Indonesian language student complaints showed that from multi 3 aspects (economics, education and health) they got 82% accuracy and 2 sentiments (positive and negative) got 80% accuracy, thus getting an average accuracy value of 81%. It can be concluded that this accuracy can be used as a multi-aspect classification and sentiment analysis.

KEYWORDS Analisis Sentiment, Covid-19, Deep Learning, Long Short Term Memory, Multi Aspect.

ABSTRAK Banyaknya data aduan Mahasiswa Universitas Muhammadiyah Sidoarjo (UMSIDA) yang terdampak wabah pandemi Covid19, dengan pemberlakuan pembatasan kegiatan masyarakat (PPKM). UMSIDA membentuk sebuah tim yang diberi nama Umsida Covid-19 Command Center (UCCC), dengan tujuan pelaksanaan program pecegahan dan aksi penanganan Covid-19, dengan harapan peneliti ingin mempermudah penyampaian informasi / aduan mahasiswa, khususnya terhadap tim UCCC sebagai bahan pertimbangan dalam melakukan suatu keputusan untuk menghadapi pandemi covid saat ini. Multi aspect sentiment analysis menghadirkan sesuatu yang baru, untuk memahami pendapat dan penilaian pengguna yang diungkapkan secara online. Dengan tujuan untuk mengklasifikasikan teks subjektif dengan memberi label polaritas, Pembentukan representasi vektor kata menggunakan Word Embedding Global Vector (Glove) dilakukan secara kombinasi dengan pelatihan analisis sentiment dengan klasifikasi berbasis Long Short Term Memory (LSTM). Pemodelan aduan mahasiswa dilakukan untuk mendapatkan representasi vektor menggunakan LSTM. Di sini, setiap kata dari kalimat menempati satu langkah pemrosesan LSTM, dan output dari kata terakhir digunakan sebagai ekspresi kalimat. Hasil dari penelitian menggunakan aduan mahasiswa bahasa Indonesia menunjukkan dari multi 3 aspect (ekonomi, pendidikan dan kesehatan) mendapatkan akurasi 82% dan 2 sentiment (positif dan negatif) mendapatkan akurasi 80% dengan demikian didapatkan nilai rata-rata Akurasi 81%. dapat disimpulkan akurasi tersebut bisa digunakan sebagai klasifikasi multi aspect dan sentiment analisis.

KATA KUNCI Analisis Sentiment, Deep Learning, Covid-19, Long Short Term Memory, Multi Aspect.

I. PENDAHULUAN

Awal tahun 2020, seluruh dunia dihebohkan dengan mewabahnya varian baru virus corona, yang diberi nama Coronavirus Disease 2019 (COVID-19). Asal dan usul virus ini diketahui dari kota Wuhan, China, pada akhir Desember 2019. Sejauh ini, 213 negara telah dikonfirmasi, termasuk Indonesia (7 Juli 2020) [1]. Di Indonesia saat ini keadaan semakin genting, dikarenakan virus Corona yang sangat cepat menyebar ke berbagai penjuru, serta banyaknya varian baru yang muncul dari virus ini, sehingga ribuan orang meninggal dunia akibat Covid. Dalam mengantisipasi penyebaran virus ini, pemerintah Indonesia telah menerapkan beberapa kebijakan, mulai dari era New Normal, PSBB (Pembatasan Sosial Berskala Besar) [2], (Pemberlakuan Pembatasan Kegiatan Masyarakat) di wilayah Jawa-Bali, PPKM Mikro, hingga PPKM Darurat [3]. Dengan berlakunya PPKM banyak sekali masyarakat atau mahasiswa yang dirugikan dalam hal perekonomian, pendidikan dan kesehatan. Masa PPKM ini berlaku kurang lebih selama 3 minggu.

Pelaksanaan PPKM ini banyak berimplikasi pada masyarakat kelas bawah dan atas dalam kaitannya dengan kebiasaan sehari-hari, mulai dari penutupan jalan, pembatasan jam malam hingga bekerja dengan WFH 50-100%. Pemberlakuan PPKM ini juga berdampak pada mahasiswa/pelajar, sebagaimana telah disampaikan oleh Menteri Pendidikan bahwa pembelajaran tatap muka yang awalnya telah disetujui dan dimulai pada bulan Juli dengan menerapkan protokol kesehatan akhirnya ditunda untuk yang kesekian kalinya.

Selain berdampak dibidang pendidikan dengan pembelajaran tatap muka yang ditunda, sebagian besar mahasiswa juga berdampak dibidang kesehatan dan ekonomi khususnya mahasiswa di Universitas Muhammadiyah Sidoarjo. Beberapa mahasiswa terdampak dalam bidang kesehatan yang terpapar Covid-19 beserta keluarga, yang mengharuskan mahasiswa dan keluarga melakukan isolasi mandiri dirumah. Dengan melakukan isolasi mandiri dirumah, secara tidak langsung anggota keluarga mahasiswa yang bekerja akan diliburkan, sehingga pemasukan ekonomi keluarga juga berkurang yang mengakibatkan mahasiswa beserta keluarga mengalami kesulitan dalam menjalani kehidupan sehari-hari. Hal ini menunjukkan bahwa pemberlakuan PPKM sangat berdampak sekali pada masyarakat, khususnya mahasiswa

Banyaknya mahasiswa Universitas Muhammadiyah Sidoarjo (UMSIDA) yang terdampak pandemi Covid 19 dengan pemberlakuan PPKM ini, Umsida membentuk sebuah tim yang diberi nama Umsida Covid-19 Command Center (UCCC) yang merupakan bagian dari MCCC (Muhammadiyah Covid-19 Comand Centre) yaitu tim pelaksanaan program pencegahan dan aksi penanganan Covid-19 [4]. UCCC ini dibentuk untuk melaksanakan program yang sama dengan MCCC, namun dalam lingkup kampus Universitas Muhammadiyah Sidoarjo.

Dari program UCCC ini ada banyak data aduan yang dapat dimanfaatkan oleh peneliti. Peneliti ingin mempermudah penyampaian informasi / aduan mahasiswa

khususnya terhadap tim Umsida Covid-19 Command Center (UCCC) sebagai bahan pertimbangan dalam melakukan suatu keputusan untuk menghadapi pandemi saat ini

Analisis opini bertujuan untuk mengklasifikasikan data tekstual menjadi pernyataan yang mendukung, bertentangan, atau menetralkan pernyataan atau tujuan [5] daripada peneliti di bidang bahasa, pemersatu pikiran atau pendapat. Pengolahan dilakukan untuk secara otomatis melakukan analisis opini pada ringkasan teks, pengumpulan informasi, pengambilan keputusan, dan survei lainnya [6], [7]. Analisis opini dapat dianggap sebagai bagian dari penambangan opini [8], dan banyak teknik algoritmik seperti K-NN, pohon keputusan, dan hutan acak dapat digunakan untuk analisis opini [9], yang merupakan hasil penelitian menggunakan LSTM. akurasi sangat tinggi dalam analisis opini pada SemEval 2016 menggunakan algoritma dan arsitektur deep learning [10], dan analisis opini publik yang baik pada berita kesehatan berbahasa Indonesia menggunakan model kalimat berbasis LSTM, mencapai akurasi tinggi [11].

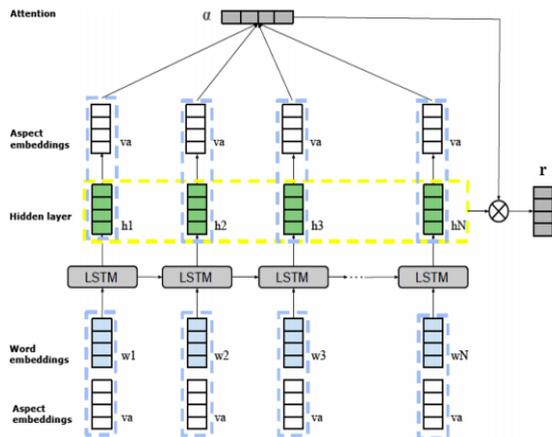
Tujuan umum dari pengolah kata hari ini adalah analisis sentimen dan klasifikasi teks [12], [13], di mana peneliti dapat menggunakan analisis sentimen berbasis aspek atau penambangan opini, yang merupakan studi opini tekstual dengan bantuan komputer. Pendapat dapat dibagi menjadi dua nilai suasana hati: positif dan negatif. Serta 3 aspect base yaitu pendidikan, ekonomi dan kesehatan. Berdasarkan konsep-konsep yang digambarkan di atas peneliti bermaksud untuk melakukan penelitian dimasa pandemi COVID-19 pada masa PPKM dengan judul penelitian "Aspect-Based Sentiment Analysis Terhadap Aduan Mahasiswa UMSIDA Di Masa Pandemi Menggunakan LSTM".

II. TINJAUAN PUSTAKA

Penelitian Mohammad Al-Smadi, Bashar, Mahmoud AlAyyoub dan Yaser Jararweh [14] melakukan analisis sentiment berbasis aspect dari ulasan Hotel Arab menggunakan dua implementasi jaringan saraf memori jangka pendek (LSTM). Yang pertama adalah (a) LSTM dua arah tingkat karakter bersama dengan pengklasifikasi bidang secara acak bersyarat (Bi-LSTM-CRF) untuk ekstraksi ekspresi target opini aspect (OTE), dan yang kedua adalah (b) LSTM berbasis aspect untuk klasifikasi polaritas sentiment aspect di mana aspectOTE dianggap sebagai ekspresi untuk mendukung identifikasi polaritas sentiment.

Misalnya, ulasan tentang Hotel mungkin berisi ekspresi sentiment yang berbeda dalam ulasan yang sama (misalnya, "Harga kamar terjangkau, tetapi lokasi hotel jauh dari pusat kota"). Review seperti contoh ini membutuhkan jenis analisis sentiment lain yang disebut analisis sentiment berbasis aspect (ABSA). ABSA memperhatikan meninjau aspect dan sentiment yang sesuai (yaitu, Harga Kamar: Positif, Lokasi Hotel: Negatif). Dengan demikian, tiga tugas utama dapat ditentukan untuk ABSA T1: identifikasi kategori aspect

alokasi, T2: ekstraksi ekspresi target opini aspect (OTE), dan T3: identifikasi polaritas sentiment aspect, pada penelitian ini berfokus pada tugas T2 dan T3 saja, Peneliti melakukan literatur LSTM dengan aspect embedding (ATAELSTM), LSTM tergantung target, dan target-koneksi LSTM. Seperti yang seperti gambar 1.



GAMBAR 1. model LSTM klasifikasi polaritas sentiment (AB-LSTMPC) [5]

LSTM berbasis aspect untuk model klasifikasi polaritas sentiment (AB-LSTM-PC) menghitung (a) apa yang disebut vektor bobot perhatian berbasis pada saling ketergantungan antara kata-kata yang dimodelkan (mis. embeddings kata) dan aspect potensinya (yaitu aspect embeddings), (b) representasi tersembunyi berbobot r yang mewakili kalimat dengan aspect tertentu Dataset disiapkan bagian dari SemEval-2016 Task-5 pada ABSA multibahasa Dataset memiliki 24.028 Tupel beranotasi ABSA dibagi sebagai (19.226) untuk pelatihan dan (4802) untuk pengujian. Selain itu, dataset telah dianotasi pada 5 tingkat teks (2291 teks ulasan) dan tingkat kalimat (6029 kalimat beranotasi). Hanya tingkat kalimat menjadi sasaran dalam penelitian ini.

Hasil eksperimen menunjukkan pendekatan yang peneliti lakukan mengungguli eksperimen dasar dan terkait pada dataset, hasil baseline sebagai berikut: ($F-1 = 31\%$) untuk T1, dan (Akurasi = 76%) untuk T2. Hasil untuk pendekatan sebagai berikut: untuk T1, Bi-LSTM-CRF berdasarkan embeddings kata word2vec tercapai ($F-1 = 66\%$) dan ($F-1 = 67\%$) di embeddings karakter faestText untuk tugas yang sama.

Pada Peneliti Analisis Sentiment berbasis aspect untuk perjalanan dan pariwisata ini mengumpulkan data ulasan hotel dan ulasan restoran di masing-masing halaman utama, status dan komentar Facebook [15]. Analisis Sentiment dapat dikategorikan menjadi 3 bagian diantara lain: sebagai analisis sentiment tingkat dokumen, analisis sentiment tingkat kalimat dan berbasis aspect analisis sentiment sebagai

berikut. Kategori dokumen: Sentiment keseluruhan dari dokumen lengkap diputuskan di level ini.

Misalnya, jika review produk diberikan, tugasnya adalah untuk memutuskan apakah itu menyampaikan keseluruhan negatif atau pendapat positif tentang produk. Tugasnya adalah memverifikasi apakah seluruh dokumen negatif, positif atau netral.

Kategori kalimat: Pekerjaan pada fase ini adalah terbatas pada kalimat dan uji apakah setiap kalimat menyampaikan pendapat negatif, positif atau netral. Pertama, kalimat diklasifikasikan sebagai tujuan atau subyektif dan kemudian kalimat yang subyektif dikategorikan sebagai positif, negatif atau netral. Kategori aspect dan entitas: kategori ini lebih menantang daripada dua lainnya.

Peneliti mengumpulkan data ulasan, postingan status, dan komentar dari halaman Facebook. Peneliti hanya mengumpulkan data bahasa Myanmar Dalam Peneliti ini kumpulan sekitar 1000 (seribu) ulasan pelanggan untuk domain hotel dan restoran dari media social Halaman Facebook.

Hasil dari jaringan (LSTM dua arah) mengambil data yang sama sebagai yang pertama dan mengembalikan prediksi yang aspect yang paling mungkin terkandung dalam kalimat dan menggunakan hasil dari jaringan pertama untuk mengembalikan jumlah aspect yang benar.

Pada peneliti ini meneliti tentang komentar wisatawan meliputi informasi pemasaran, umpan balik ke layanan dan sentiment pelanggan, yang sangat penting untuk meningkatkan pengelolaan situs wisata dengan cara: merancang pola operasi yang lebih efektif.

Analisis sentiment untuk komentar turis adalah kombinasi pemrosesan bahasa alami dan pembelajaran mesin. Berbeda dengan bahasa Inggris tersegmentasi pada spasi, teks Cina membutuhkan tokenizer seperti sebagai Jieba dan ICTCLAS dll untuk mengubahnya menjadi urutan kata-kata.

Penyematan kata yang memetakan kata atau frasa ke vector bilangan real banyak digunakan untuk menangkap sintaksis yang bermakna dan keteraturan semantik membantu algoritma pembelajaran untuk mencapai kinerja yang lebih baik dalam tugas pemrosesan bahasa alami dengan mengelompokkan kata-kata yang mirip.

Basis data peneliti ini berisi 1.000.000 komentar berlabel tiga kelas, positif, netral dan negatif dilambangkan sebagai 1, 2 dan 3 masing-masing. Panjangnya komentar bervariasi dari nol hingga 536 karakter Cina, umum dalam 35 hingga 155 karakter Cina. Panjang rata-rata komentar adalah sekitar 76 karakter Cina.

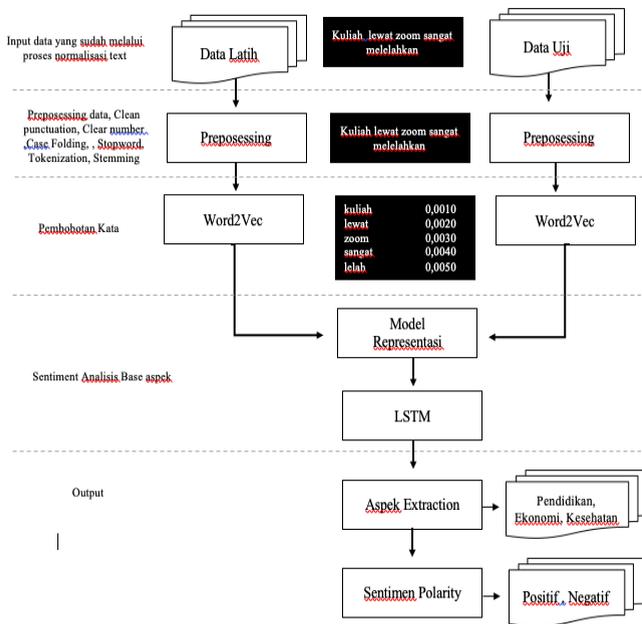
Hasil eksperimen menunjukkan bahwa akurasi untuk analisis sentiment dengan CNN yang dibangun di atas model

LSTM ditingkatkan masing-masing sebesar 3,13% dan 1,71%, dibandingkan dengan yang terpisah CNN dan model LSTM.

III. PERANCANGAN

Dalam pembahasan ini, peneliti akan mengelaborasi desain arsitektur sistem, serta arsitektur dan algoritme alternatif yang digunakan.

Peneliti membuat desain arsitektur multi aspect dan sentiment dengan urutan sebagai berikut : pertama data aduan mahasiswa, kedua proses untuk preprocessing data aduan agar nanti saat di proses ke LSTM data lebih baik, ketiga proses word embedding saat ini peneliti menggunakan global vector (Glove) dalam penanganannya, keempat proses perhitungan menggunakan metode long short term memory (LSTM), kelima yaitu mencari hasil dari aspect based dan keenam mencari hasil sentiment analisis, seperti itu konsep dalam penelitian ini, arsitektur penelitian klasifikasi multi aspect dan sentiment ditunjukkan pada gambar. 2

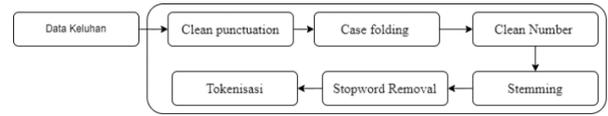


GAMBAR 2. Arsitektur Penelitian Multi Aspect dan Sentiment Analisis
Gambar, 4 Arsitektur Penelitian Multi Aspect dan Sentiment Analisis

A. PREPROCESSING

Proses preprocessing ini adalah proses yang berjalan sebelum mengolah data dengan cara yang ingin digunakan. Proses ini berjalan secara sistematis dalam sebuah aplikasi pengolah data Python. Penggantian Singkatan Proses ini merupakan perubahan singkatan yang diubah menjadi teks. Singkatan dan kata-kata sederhana diterjemahkan ke dalam bahasa Indonesia yang baik oleh tim analis khususnya tim

umsida covid command centre (UCCC) yang bergerak di bidang kesosialan.



GAMBAR 3. Tahap Preprocessing

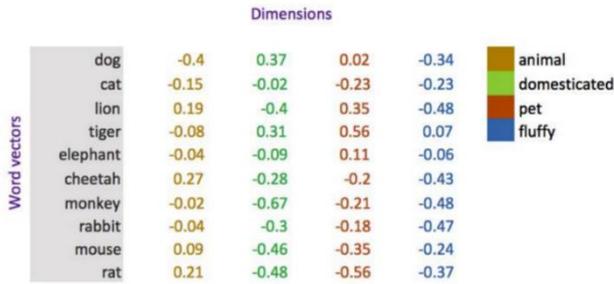
Gambar 3, 6 tahap preprocessing:

- Clean punctuation**
yaitu Menghapus tanda baca yang ada dalam dokumen kecuali tanda koma (,) atau titik(.) agar bisa memisahkan bebrapa kalimat.
- Case Folding**
Case folding adalah Melipat huruf besar dan huruf kecil adalah proses mengubah semua huruf besar menjadi huruf kecil.
- Clean Number**
Clean Number jika menggunakan angka sebelum atau sesudah kata dalam data aduan mahasiswa anda, angka tersebut akan dihapus karena tidak ada manfaat dari survei ini.
- Stemming**
Steeming yaitu proses menghilangkan awalan, sisipan, dan akhiran. Sehingga kata yang disimpan atau diproses adalah sebuah kata dasar.
- Stopword Removal**
menghapus stopwords dalam proses ini menghapus kata-kata yang sering muncul tetapi tidak berpengaruh. Kata-kata yang mengandung kata ini adalah kata-kata seperti "yang", "dan", "di", "dari", termasuk protokol kesehatan dan virus corona/covid19..
- Tokenisasi**
Tokenisasi ini adalah proses memisahkan kata-kata. Fase tokenizer dimulai dengan memisahkan bagian-bagian aduan yang dipisahkan oleh ruang. Kemudian token diberikan pada beberapa kata yang diproses pada tahap selanjutnya

B. PROSES WORD EMBEDDING

Proses penyisipan kata ini melakukan teknik yang mengubah kalimat teks menjadi teks vektor yang mewakili makna kata. Ukuran vektor bervariasi dari 50 hingga 300 vektor dimensi [16]. Nilai vektor ini nantinya akan digunakan sebagai masukan untuk proses pelatihan dan pengujian LSTM. Setiap catatan pengaduan mahasiswa dihitung dengan catatan terpanjang dalam catatan. Kemudian, jika kalimat lebih pendek dari panjang kalimat terpanjang, sisa kalimat diberi skor 0, atau kata-kata acak yang tidak dipahami diberi skor 0. Atau, umumnya dikenal sebagai padding dengan menambahkan 0 sebelum dan

sesudah himpunan untuk menghitung setiap himpunan dengan panjang vektor yang sama. Untuk mengubah kalimat menjadi vektor, penulis menggunakan GloVe (Global Vector) [17]. Berikut adalah contoh kata vektor



GAMBAR 4. WordVector[15]

Setiap kata dalam vektor memiliki nilai yang sama dengan dimensi yang ditentukan. Nilai vektor kata sebuah kata mungkin memiliki kesamaan dengan kata-kata serupa. Misalnya, singa dan cheetah memiliki statistik yang tidak jauh dari dimensi hewan.

C. Proses Training LSTM

Dalam proses pelatihan LSTM ini, peneliti mempelajari aplikasi RNN. RNN adalah sejenis jaringan syaraf tiruan, dan output yang dihasilkan dapat digunakan kembali sebagai proses input untuk perhitungan lebih lanjut. Hal ini menyebabkan RNNs memperhatikan urutan sebagai salah satu karakteristik pembelajaran [18].

RNN juga dapat disembunyikan. Ini adalah memori jaringan saraf yang dihitung dari keadaan tersembunyi sebelumnya. Long short-term memory (LSTM) adalah jenis RNN dengan empat lapisan jaringan saraf tiruan yang berinteraksi di setiap keadaan tersembunyi [8]. LSTM juga memiliki empat jenis gerbang yang berinteraksi dan berada dalam keadaan sel. Hal ini memungkinkan informasi mengalir dari satu keadaan tersembunyi ke keadaan berikutnya, tanpa banyak atau tanpa perubahan [19].

Dalam arsitektur RNN, ada input X, keadaan tersembunyi A, dan output t, masing-masing pada waktu t. Keadaan tersembunyi Di dalam (1) adalah memori jaringan syaraf tiruan yang dihitung dari keadaan tersembunyi sebelumnya. U, V, dan W adalah bobot beberapa lapisan RNN. Tidak seperti ANN biasa yang memiliki bobot berbeda untuk setiap layer, metode RNN memiliki bobot yang sama untuk semua layer.

$$A_t = f(x_t.U + A_{t-1}.W) \quad (1)$$

Peneliti dapat menggunakan fungsi tanh atau ReLU untuk fungsi f. Hipotesis h diberikan pada (2).

$$h_t = \text{softmax}(A_t.V). \quad (2)$$

Langkah pertama dalam mengoperasikan jaringan LSTM adalah menggunakan gerbang oblivion f yang

dijelaskan pada (3) untuk menghapus informasi dari status sel. Vektor pertama yang disalin 1 dilewatkan ke gerbang oblivion terlebih dahulu. Dimana b adalah nilai bias.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f). \quad (3)$$

Jaringan LSTM kemudian menggunakan gerbang input yang ditunjukkan pada (4) untuk memperbarui informasi baru tentang status sel. Gerbang ini bertanggung jawab untuk menentukan informasi tentang status sel yang akan diperbarui.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i). \quad (4)$$

Pada saat yang sama, vektor yang sama dilewatkan ke lapisan tanh, rentang nilai dalam vektor diubah menjadi bilangan real antara -1 dan 1, dan keadaan et dari sel kandidat yang ditunjukkan pada (5).

$$\tilde{C}_t = \text{tanh}(W_c.[h_{t-1}, x_t] + b_c). \quad (5)$$

Vektor yang melewati gerbang input dan lapisan tanh kemudian melakukan operasi perkalian titik demi titik, hanya menyisakan output yang diinginkan dalam keadaan sel. Dengan demikian, manipulasi vektor setelah melewati tiga gerbang LSTM dan lapisan tanah adalah untuk menghasilkan nilai keadaan sel yang sesuai (6).

$$C_t = f_t.C_{t-1} + i_t.C_t. \quad (6)$$

Langkah terakhir adalah menentukan gerbang keluaran pada (7). Vektor 1 disalin dan diteruskan ke gerbang keluaran.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o). \quad (7)$$

Pada saat yang sama, itu disalin dan diteruskan ke Lapistan. Setelah melewati lapisan tanh, ia bertemu Ot dalam operasi perkalian pointwise dan menjadi nilai output dari sel yang sesuai (8).

$$h_t = o_t * \text{tanh}(C_t). \quad (8)$$

LSTM memiliki gerbang lupa gabung dan varian lain yang menggunakan gerbang input yang disebut gerbang lupa input gabung LSTM. Variasi ini menentukan ukuran gerbang pelupa dan gerbang input pada saat yang bersamaan, tidak terpisah. Dalam variasi ini, ukuran gerbang input tergantung pada ukuran gerbang oblivion yang ditunjukkan pada (9).

$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t. \quad (9)$$

Dalam variasi ini, ukuran gerbang input tergantung pada ukuran gerbang oblivion. Rumus diatas merupakan rumus untuk fase forward propagation LSTM yang digunakan untuk menerima input dan menghasilkan output data pengaduan. Fase Forward Propagation Setelah proses, ada juga fase backpropagation yang membantu menghitung

jumlah perubahan yang diperlukan untuk memperbarui bobot yang ada. Berbeda dengan fase forward propagation, proses pada fase backpropagation berjalan dari last time step ke first time step.

Analisis opini menggunakan LSTM menggunakan LSTM untuk memproses pernyataan pada setiap langkah waktu [20]. Analisis sentimen dan tugas berbasis aspek ketidakpuasan mahasiswa menggunakan representasi vektor ketidakpuasan mahasiswa untuk Covid19 yang diberikan sebagai input ke LSTM pada langkah pertama, dan representasi vektor terkait ketidakpuasan mahasiswa adalah yang kedua. LSTM.

Varian LSTM terutama digunakan untuk analisis dan multi-aspek. Yaitu, Coupled-Input-Lupakan-Gate LSTM. Keluaran dari langkah kedua adalah representasi pengaduan mahasiswa terhadap data pengaduan yang bersangkutan. Outputnya kemudian diteruskan ke lapisan yang terhubung penuh untuk mendapatkan probabilitas label. Label probabilitas kemudian diteruskan ke lapisan softmax untuk analisis sentimen dan ke lapisan sigmoid untuk basis multi-aspek untuk probabilitas yang lebih baik.

IV. PENGUJIAN DAN ANALISIS

A. DATASET

Adapun dataset yang digunakan pada penelitian ini adalah data aduan mahasiswa umsida sebanyak 3000 dokumen. Dokumen tersebut bersumber dari penyebaran kuesioner pada bulan April-Juli 2021 yang di ambil secara online. Pada penelitian ini pengambilan data mahasiswa tersebut menggunakan sistem umsida berbasis website dengan url <https://form.umsida.ac.id>.

Setelah penyebaran kuesioner data aduan hasil dikumpulkan jadi satu bentuk excel. Setiap aduan akan diberikan label secara manual yang mewakili keluhan mahasiswa, label kelas multi terdiri dari 3 jenis yaitu pendidikan, kesehatan, dan ekonomi yang di jadikan multi label menjadi 7 kelas, serta terdapat sentiment analysis yaitu positif dan negatif.

Dengan adanya data aduan dari mahasiswa peneliti meminta bantuan dari team UCCC untuk memberikan pelabelan aspect based dan sentiment analysis supaya data aduan mahasiswa terlabel dengan baik dan benar, serta membetulkan kata – kata yang typo atau singkatan yang bisa berpengaruh terhadap dataset. Setelah terlabel semua peneliti menghitung jumlah label data aduan mahasiswa yang sudah terkumpul dan dapat dilihat pada Tabel I.

TABEL I
 Jumlah Data Aduan Sudah Terlabel

Aspect	Jumlah	Positif	Negatif
Ekonomi	500	250	250
Pendidikan	350	150	100
Kesehatan	600	200	400

Ekonomi Pendidikan	450	300	150
Ekonomi Kesehatan	325	125	200
Pendidikan Kesehatan	425	325	100
Ekonomi Pendidikan Kesehatan	350	250	100

Sebagian data aduan mahasiswa yang diperoleh dari form.umsida.ac.id sistem umsida yang sudah terlabel multi aspect base dan sentiment analysis oleh peneliti dapat dilihat pada Tabel II dibawah ini.

TABEL II
 Contoh Sebagian Data Aduan Mahasiswa

Aduan Mahasiswa	Aspect	Sentiment
Saya dirumahkan dari pekerjaan, dan belum mendapatkan pekerjaan lagi	Ekonomi	Negatif
Alhamdulillah dengan adanya bantuan keringanan biaya kuliah ini ditengah pandemi covid 19 dapat membantu meringankan beban orangtua saya.	Ekonomi, Pendidikan	Positif
Konsep belajar yang diterapkan bu Ratih sudah baik, bisa ditingkatkan lagi bila ada praktik jika pandemi sudah berakhir	Pendidikan	Positif
efek vaksinasi yang dapat membuat imunitas tubuh menurun seperti terjadi gejala demam, mual mual, dan sebagainya	Kesehatan	Negatif
Saya senang sekali diajari beliau karena mengkodekan itu seru dan semoga tetap dipertahankan memberi kasus penyakit dan di bahas secara bersama-sama, sehat selalu ya Bu anna serta dilancarkan rezekinya.	Ekonomi Kesehatan	Positif
Semoga semua dosen bisa memberi toleransi kepada mahasiswa yang bekerja sambil kuliah di era pandemi banyak mahasiswa yg kesulitan ekonomi untuk membantu keluarga dengan bekerja	Ekonomi pendidikan kesehatan	Positif

B. PENGUJIAN

Pengujian dilakukan dengan Anaconda – Jupiter

1. Preprocessing

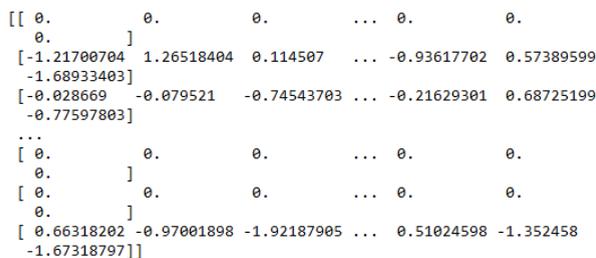
Preprocessing ini memiliki beberapa tahapan yang dilakukan oleh preprocessing setelah akuisisi data selesai diproses. Tahap pertama dari preprocessing adalah menangkap karakteristik aduan masing-masing mahasiswa. Fitur yang diwariskan meliputi huruf besar/kecil, kemampuan menulis semua kata/frasa dalam huruf kecil, kemampuan memiliki piktoqram, dan ada tidaknya kata negatif. Fitur dari setiap set disimpan dalam format vektor. Setelah karakteristik masing-masing set diperoleh, dilakukan pembersihan. Non-abjad dan angka seperti tanda tanya, titik dua, dan koma dihilangkan. Selain itu, dilakukan proses penangkapan kata dasar yang disebut stemming, yang

kemudian diolah menjadi tokenisasi setiap kata dalam kalimat dan kembali ke kata dasar.

Akuisisi root dilakukan dengan menghilangkan imbuhan depan dan belakang secara manual yang dikembangkan berdasarkan penelitian sebelumnya [21]. Daftar kata dibuat dari kata-kata yang telah mengalami pemrosesan batang. Setiap kata. Proses pemodelan kalimat kemudian memberikan setiap kalimat nomor unik.

Dalam daftar pernyataan, setiap pernyataan adalah unik (keduanya tidak sama). Oleh karena itu, proses pengungkapan kata dalam bentuk vektor dilakukan bersamaan dengan pelatihan model analitik dan disebut sebagai glove penyisipan kata terlatih [22]. Nilai vektor kata dimasukkan secara acak dan dilatih menggunakan teknik jaringan saraf tiruan yang termasuk dalam arsitektur LSTM.

Word Embedding atau Penyisipan kata adalah pemodelan bahasa dan teknik pembelajaran fungsional di bidang pemrosesan bahasa alami. Setiap kata atau frasa dalam kosakata dikaitkan dengan vektor yang berisi angka. Di lapisan tertanam, nilai vektor ditentukan oleh tipe data nyata dengan panjang dan nilai tertentu untuk setiap vektor dalam kamus. Panjang vektor adalah salah satu parameter pelatihan yang dimodifikasi untuk tujuan percobaan ini. Parameter ini digunakan untuk menentukan ukuran embed. Pertama, nilai acak diberikan ke vektor untuk setiap ID kata. Nilai setiap vektor berubah seiring berjalannya proses pelatihan. Pelatihan dilakukan dengan menggunakan jaringan syaraf tiruan. Setelah dilakukan pelatihan, kata-kata yang memiliki kesamaan makna memiliki vektor-vektor yang nilainya saling berdekatan, seperti terlihat pada Gambar 5 di bawah ini.



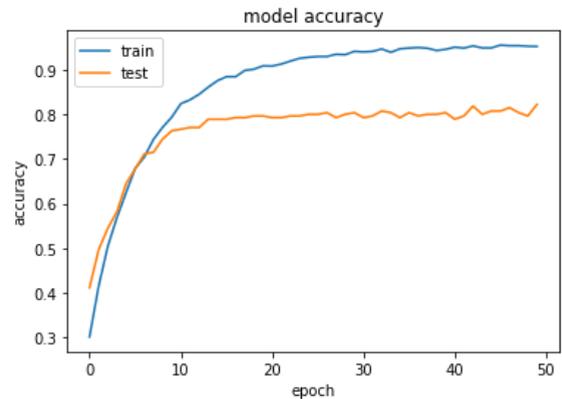
GAMBAR 5. Vector Vector dalam satuan kalimat

Dalam gambar 5 peneliti menggunakan 50 dimensi untuk rumus dalam menggunakan word embedding global vector atau di sebut Glove.

2. Pengujian Model

Dalam pengujian, kami menguji model jaringan saraf tiruan ini menggunakan metode LSTM dan parameter LSTM untuk menemukan arsitektur yang optimal. Semua eksperimen dilakukan dengan data uji 5% hingga 20%, total 3000 data

aduan mahasiswa yang relevan, hingga 50 hingga 100 epoch, dan tingkat pembelajaran 0,01 hingga 0,02.

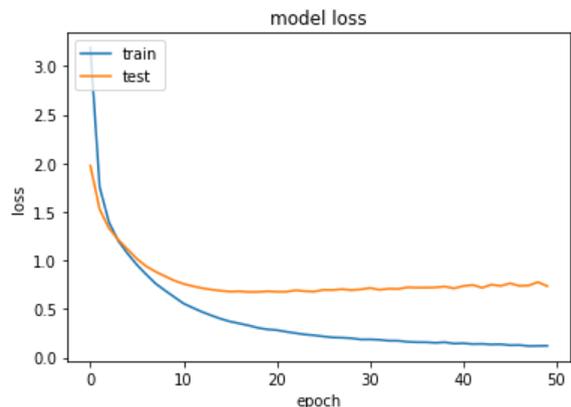


GAMBAR 6. Grafik Validasi Akurasi Multi Aspek

Uji model dilakukan pada dataset Indonesia dengan tujuh aspek: ekonomi, pendidikan kesehatan, ekonomi kesehatan, ekonomi pendidikan, kesehatan pendidikan, dan ekonomi kesehatan pendidikan. Percobaan ini meliputi pengujian pengaruh metode word embedding glove, dimensi embedding (ukuran embedding) 50D-150D, pada metode pemodelan kalimat, metode klasifikasi, ukuran keluaran LSTM, dropout, dan fitur tambahan lainnya.

Data hasil Word Embedding Glove akan dijalankan untuk training dengan metode LSTM, data tersebut akan divalidasi dengan skor f1, fit and recall, dan anda akan dapat menentukan hasil akurasi dari data test dan training.

Hasil membuktikan hasil dari aspect base adalah: Akurasi 82%, kemudian hasil dari Sentiment Analisis mendapatkan Akurasi 80% dari hasil kedua aspect dan sentiment ini peneliti rata-rata mendapatkan hasil 81%. Berikut adalah latihan data untuk mengetahui validasi akurasi dari dataset gambar 6. Untuk multi aspect based.



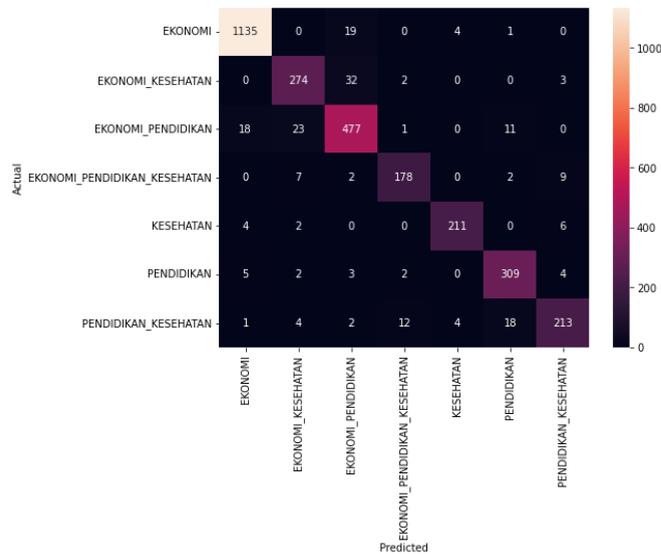
GAMBAR 7. Grafik validation Loss Multi Aspect

Hasil grafik dari *validation akurasi* saat epoch ke 7 ada titik temu data train dan data test, kemudian saat melaju ke epoch 8 mulai terjadi pemisahan antara data test dan data train, dari epoch 10 – 50 data train mengalami peningkatan sedangkan data test stabil di akurasi 80%.

Selanjutnya, tingkat kehilangan verifikasi karena penyematan kata GloVe dan LSTM berbasis aspek dapat dikonfirmasi dalam grafik berikut di mana eksperimen yang sama dilakukan Gambar 7.

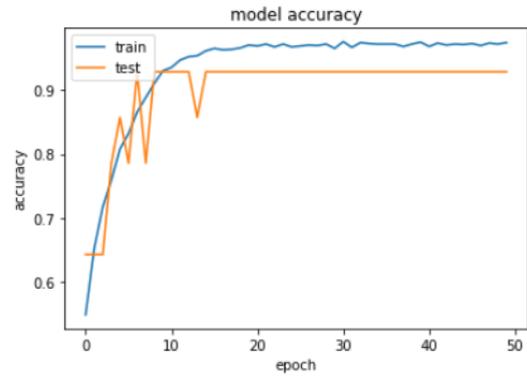
Hasil grafik dari *validation loss* multi aspect pada epoch ke 5 terdapat titik temu data test dan train, akan tetapi dalam epoch selanjutnya data train semakin menurun dan data test stabil di 10%.

Setelah medapat hasil akurasi aspect base peneliti membuat tabel matrik untuk pengecekan dari hasil training dan testing hasil aktual dan prediksi ekonomi: 1135, pendidikan: 309, kesehatan : 211, ekonomi pendidikan : 477, ekonomi kesehatan : 274, pendidikan kesehatan : 213, ekonomi pendidikan kesehatan : 178 jadi ada sebagian data aduan mahasiswa yang tidak sesuai dengan label manual dan prediksi sistem yang ditampilkan seperti gambar 8 di bawah ini.



GAMBAR 8. Matriks Actual & Prediksi Aspect Base

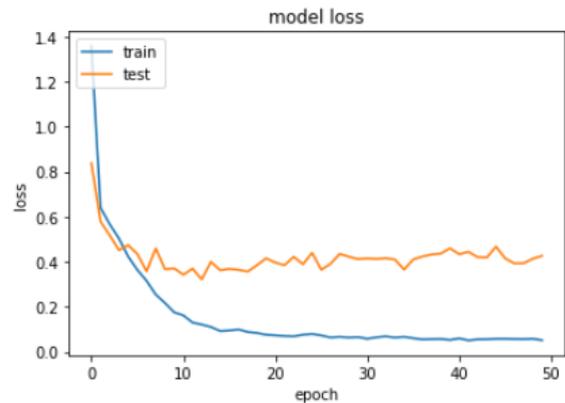
Setelah mendapatkan hasil training dan testing dari aspect based, kemudian di ulang kembali ke proses semula, setelah dataset aduan mahasiswa ditraining dan juga ditesting ke sentiment analysis, supaya pada mendapat model sesuai dengan sentiment analisis. Berikut hasil grafik validasi akurasi sentiment analisis gambar 9.



GAMBAR 9. Ilustrasi validasi Akurasi Sentiment Analisis

Hasil grafik *validasi akurasi* sentiment analisis pada data training saat epoch 10 sampai 50 sudah stabil yang di tunjukkan dengan garis warna biru kemudian meningkat secara perlahan, sedangkan data testing mulai epoch 7 sampai epoch 15 nilai akurasi naik dan turun setelah epoch 16 grafik sudah membaik.

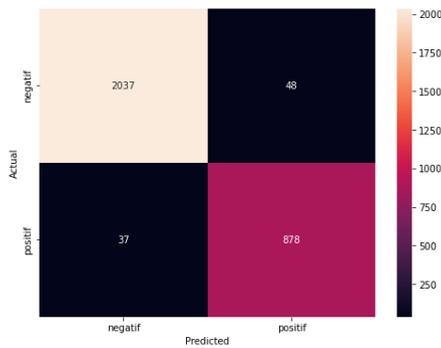
Selanjutnya, grafik berikut menunjukkan tingkat kehilangan validasi karena penyisipan kata GloVe dan LSTM untuk analisis sentimen dalam eksperimen yang sama pada Gambar 10.



GAMBAR 10. Ilustrasi Validasi Loss Sentiment Analisis

Dari hasil *validasi loss* data training mengalami kenaikan pada epoch ke 10 dan data testing naik turun pada epoch ke 7 samapi epoch 50.

Setelah medapat hasil akurasi aspect based peneliti membuat tabel matrik untuk pengecekan dari hasil training dan testing hasil aktual dan prediksi negatif yang sesuai mendapat 2037 kalimat, sedangkan yang aktual dan prediksi positif mendapat 878 kalimat seperti matrik gambar 11 di bawah ini.



GAMBAR 11. Matriks Sentiment Analisis

Dengan demikian hasil yang di peroleh dari multi aspect dan sentiment analysis berbeda, akan tetapi model ini bisa melakukan pemrosesan degan baik

V. KESIMPULAN

Implementasi berbasis multi aspek menggunakan metode LSTM dan analisis uji analisis sentiment, Kesimpulan dari hasil analisis dapat digunakan untuk menyimpulkan bagaimana mengambil keputusan menggunakan kumpulan data pengaduan mahasiswa dari UCCC umsida Covid Command Center dapat disimpulkan bisa. Pemrosesan covid19 yang lebih cepat di umsida dapat dijadikan sebagai tolok ukur untuk mengevaluasi pelayanan kepada mahasiswa umsida.

Hasil evaluasi data aduan mahasiswa mendapatkan hasil akhir dengan akurasi 81%. Oleh karena itu, mengingat keakuratan penelitian ini, kami dapat menyimpulkan bahwa data yang disajikan dapat digunakan sebagai tolok ukur tim Umsida Covid Command Center (UCCC) dalam pengambilan keputusan saat menangani Covid 19. Untuk memastikan bahwa mahasiswa terus menerima layanan yang tepat dan terbaik dari tim UCCC.

Untuk Penelitian yang akan datang ketika data aduan memiliki banyak bahasa seperti bahasa indonesia, bahasa jawa, dan banyak lainnya dan juga banyak data mengandung singkatan – singkatan, data ini harus benar - benar di perhatikan di tahap preprocessingnya. Karna peneliti mengalami hal tersebut sehingga membuat hasil akhir tidak sesuai dengan harapan, juga data aduan mahasiswa harus lebih banyak dari peneliti agar LSTM bisa bekerja lebih optimal lagi.

VI. REFERENSI

- [1] W. H. Organization and others, "WHO coronavirus disease (COVID-19) dashboard. 2020." 2020.
- [2] "Peraturan Menteri Kesehatan Republik Indonesia Nomor 9 Tahun 2020 Tentang Pedoman Pembatasan Sosial Berskala Besar dalam Rangka Percepatan Penanganan Corona Virus Disease 2019 (Covid-19)." 2020. [Online]. Available: http://hukor.kemkes.go.id/uploads/produk_hukum/PMK_No_9

- [3] "_Th_2020_ttg_Pedoman_Pembatasan_Sosial_Berskala_Besar_Dalam_Penanganan_COVID-19.pdf.
- [4] "Mulai 3 Juli, Pemerintah Berlakukan PPKM Darurat di Jawa-Bali," 2021. <https://setkab.go.id/mulai-3-juli-pemerintah-berlakukan-ppkm-darurat-di-jawa-bali/> (accessed Jul. 01, 2021).
- [5] M. Al-Smadi, B. Talafha, M. Al-Ayyoub, and Y. Jararweh, "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 8, pp. 2163–2175, 2019.
- [6] S. M. Mohammad, P. Sobhani, and S. Kiritchenko, "Stance and sentiment in tweets," *ACM Trans. Internet Technol.*, vol. 17, no. 3, pp. 1–23, 2017.
- [7] J. Ebrahimi, D. Dou, and D. Lowd, "A joint sentiment-target-stance model for stance classification in tweets," in *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*, 2016, pp. 2656–2665.
- [8] P. Sobhani, S. Mohammad, and S. Kiritchenko, "Detecting stance in tweets and analyzing its interaction with sentiment," in *Proceedings of the fifth joint conference on lexical and computational semantics*, 2016, pp. 159–169.
- [9] P. Krejzl, B. Hourová, and J. Steinberger, "Stance detection in online discussions," *arXiv Prepr. arXiv1701.00504*, 2017.
- [10] A. Aker, L. Derczynski, and K. Bontcheva, "Simple open stance classification for rumour analysis," *arXiv Prepr. arXiv1708.05286*, 2017.
- [11] G. Zarrella and A. Marsh, "Mitre at semeval-2016 task 6: Transfer learning for stance detection," *arXiv Prepr. arXiv1606.03784*, 2016.
- [12] A. Zaini, M. A. Muslim, and others, "Pengelompokan artikel berbahasa indonesia berdasarkan struktur laten menggunakan pendekatan self organizing map," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 3, pp. 259–267, 2017.
- [13] J. Santoso *et al.*, "Self-Training Naive Bayes Berbasis Word2Vec untuk Kategorisasi Berita Bahasa Indonesia," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 7, no. 2, pp. 158–166, 2018.
- [14] E. I. Setiawan *et al.*, "Analisis Pendapat Masyarakat terhadap Berita Kesehatan Indonesia menggunakan Pemodelan Kalimat berbasis LSTM (Indonesian Stance Analysis of Healthcare News using Sentence Embedding Based on LSTM)," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 1, pp. 8–17, 2020.
- [15] S. Y. Maw and M. A. Khine, "Aspect based Sentiment Analysis for travel and tourism in Myanmar Language using LSTM," *MERAL Portal*, 2019.
- [16] J. Gao, R. Yao, H. Lai, and T.-C. Chang, "Sentiment analysis with CNNs built on LSTM on tourists comments," in *2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, 2019, pp. 108–111.
- [17] J. B. Ahire, "Introduction to Word Vectors," *Retrieved March*, vol. 12, p. 2018, 2018.
- [18] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [19] S. Hochreiter and J. Uergen Schmidhuber, "LONG SHORT-TERM MEMORY," *Neural Comput.*, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [20] E. Lim, E. I. Setiawan, and J. Santoso, "Stance Classification Post Kesehatan di Media Sosial Dengan FastText Embedding dan Deep Learning," *J. Intell. Syst. Comput.*, vol. 1, no. 2, pp. 65–73, 2019.
- [21] E. Kochkina, M. Liakata, and I. Augenstein, "Turing at semeval-2017 task 8: Sequential approach to rumour stance classification with branch-lstm," *arXiv Prepr. arXiv1704.07221*, 2017.

- [21] A. D. Tahitoe and D. Purwitasari, "Implementasi modifikasi enhanced confix stripping stemmer untuk bahasa indonesia dengan metode corpus based stemming," *J. Ilm*, vol. 12, no. 15, pp. 1–15, 2010.
- [22] R. Johnson and T. Zhang, "Semi-supervised convolutional neural networks for text categorization via region embedding," *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.