

SENTIMENT SUMMARIZATION EVALUASI PEMBELAJARAN MENGGUNAKAN ALGORITMA *LSTM* (LONG SHORT TERM MEMORY)

Achmad Yogie Setiawan¹, I Gede Mahendra Darmawiguna², Gede Aditra Pradnyana³

Program Studi Pendidikan Teknik Informatika
Universitas Pendidikan Ganesha
Singaraja, Indonesia

ayogiesetiawan@gmail.com¹, mahendra.darmawiguna@undiksha.ac.id², gede.aditra@undiksha.ac.id³

Abstrak— Evaluasi pembelajaran dosen adalah sebuah teks yang berisi ulasan mahasiswa terkait kinerja pembelajaran dosen. Evaluasi pembelajaran digunakan sebagai bahan refleksi diri dosen agar dapat meningkatkan pelayanan pembelajaran yang diberikan pada pembelajaran selanjutnya. Evaluasi tersebut berjumlah banyak membuat dosen kesulitan untuk menganalisis. Diperlukan teknik analisis sentimen untuk mengklasifikasikan evaluasi mahasiswa. Evaluasi yang sudah terklasifikasi masih menyisahkan teks yang panjang dan berbelit. Peringkasan teks merupakan salah satu solusi untuk meringkas teks yang panjang menjadi sebuah teks yang padat dan informatif. Peringkasan teks bermanfaat untuk menghemat waktu dalam mencari inti yang terdapat pada teks. Terdapat dua metode dalam peringkasan teks yaitu metode ekstraktif dan abstraktif. Pada penelitian ini menggunakan metode abstraktif dikarenakan data yang digunakan adalah evaluasi pembelajaran dosen yang ulasannya ditulis oleh mahasiswa. Algoritma yang digunakan untuk pengklasifikasian sentimen dan peringkasan teks ini menggunakan algoritma *Long Short Term Memory (LSTM)*. Hasil klasifikasi sentimen akan dievaluasi dengan menggunakan *confusion matrix* yaitu menguji model dengan data evaluasi. Sedangkan hasil ringkasan akan dievaluasi dengan menggunakan ROUGE yaitu membandingkan hasil ringkasan dari sistem dengan ringkasan manual oleh ahli. Pada pengujian *confusion matrix* sistem mendapatkan nilai *accuracy* sebesar 0,902 dan nilai *f-measure* sebesar 0,921. Pada pengujian *Recall-Oriented Understudy for Gisting Evaluation (ROGUE)* evaluasi positif mendapatkan nilai 0,16 dan pada evaluasi negatif mendapatkan nilai 0,2. *Tokenizer* yang dikembangkan belum dapat menyimpan token hasil proses *training*. Akibatnya hasil prediksi saat *load model* tidak akan sebagai hasil prediksi saat selesai *training*.

Kata kunci : Analisis Sentimen, Peringkasan, LSTM, ROGUE

Abstract— *Lecturer learning evaluation is a text that contains student reviews related to lecturer learning performance. Learning evaluation is used as a lecturer's self-reflection material to improve the learning services provided in the next lesson. The evaluations are many in number, making it difficult for lecturers to analyze. Sentiment analysis techniques are needed to classify student evaluations. The evaluation that has been classified still leaves a long and convoluted text. Text summarization is one solution to summarize a long text into a dense and informative text. Text summarization is useful to save time in searching for the gist of the text. There are two methods in text summarization, namely extractive and abstractive methods. This study uses an abstract method because the data used is an evaluation of lecturer learning whose reviews are written by students. The algorithm used for sentiment classification and text summarization uses the Long Short Term Memory (LSTM) algorithm. The results of the sentiment classification will be evaluated using a confusion matrix, namely testing the model with evaluation data. While the summary results will be evaluated using ROUGE, which is comparing the summary results from the system with a manual summary by experts. In testing the confusion matrix system, the accuracy value is 0.902 and the f-measure value is 0.921. In the Recall-Oriented Understudy for Gisting Evaluation (ROGUE) test, the positive evaluation scored 0.16 and the negative evaluation scored 0.2. The developed tokenizer has not been able to store the tokens resulting from the training process. As a result, the prediction results when loading the model will not be as good as the prediction results when finished training.*



KARMAPATI

Kampus Akademik Riset dan Inovasi

Keywords : *Sentiment Analysis, Summarization, LSTM, ROGUE*

e-ISSN: 2685-7006 | p-ISSN: 2252-9063

*Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika
(KARMAPATI)*

Volume 11, Nomor 2, Tahun 2022

I. PENDAHULUAN

Informasi merupakan salah satu kebutuhan penting yang tidak dapat terlepas dari kehidupan manusia. Membaca menjadi salah satu cara manusia memperoleh informasi, baik itu membaca buku, , berita, majalah, artikel, jurnal, dan lain-lain. Pada perguruan tinggi untuk meningkatkan kinerja pembelajaran seorang dosen membutuhkan sebuah evaluasi pembelajaran untuk mengetahui pengalaman belajar mahasiswa pada mata kuliah tersebut. Evaluasi pembelajaran digunakan sebagai bahan refleksi diri dosen agar dapat meningkatkan pelayanan pembelajaran yang diberikan pada pembelajaran selanjutnya.

Universitas Pendidikan Ganesha (Undiksha) menyediakan angket online untuk setiap mata kuliah melalui Sistem Informasi Akademik yang dapat diisi oleh mahasiswa pada setiap akhir semester. Angket *online* yang disediakan pada SIAK Undiksha berupa isian responden dalam bentuk kalimat kritik dan saran. Pada angket tersebut, mahasiswa mencurahkan pengalamannya selama diajar oleh dosen pada mata kuliah yang diajarkan, atau disebut juga evaluasi pembelajaran dosen. Evaluasi pembelajaran dosen adalah sebuah teks yang berisi ulasan mahasiswa terkait kinerja pembelajaran dosen.

Banyaknya jumlah mahasiswa yang diajar oleh dosen pada setiap mata kuliah membuat dosen meluangkan waktu yang lebih untuk membaca setiap evaluasi yang ada. Evaluasi yang berisi ulasan kritik dan saran membuat dosen kesulitan untuk menganalisis evaluasi apakah mahasiswa merasa puas atau tidak dengan pelayanan pembelajaran yang diberikan. Untuk mengefisienkan waktu dosen, maka dibutuhkan yaitu sistem yang dapat membantu dosen dalam melakukan proses analisis evaluasi pembelajaran secara otomatis. Evaluasi pembelajaran dosen diklasifikasikan kedalam bentuk evaluasi positif dan negatif, pengklasifikasian tersebut disebut *opinion mining* atau juga disebut analisis sentimen.

Analisis Sentimen sudah banyak dimanfaatkan oleh penyedia jasa atau produk sebagai bahan *review* terhadap pelayanan atau produk yang mereka tawarkan. Analisis sentiment dilakukan dengan cara pengklasifikasian *review* menjadi positif dan negatif. Evaluasi pembelajaran yang sudah terklasifikasi kedalam evaluasi positif dan negatif masih menyisahkan teks yang panjang dan berbelit. Teks evaluasi tersebut membuat dosen membutuhkan waktu yang lebih untuk mencari intisari yang ada. Oleh karena itu pada beberapa jenis teks terdapat bagian ringkasan untuk mempersingkat waktu dalam mencari intisari teks. Ringkasan merupakan cara efektif untuk menyajikan karangan yang panjang menjadi bentuk singkat. Evaluasi pembelajaran dosen merupakan salah satu jenis teks yang tidak memiliki ringkasan.

Oleh karena itu dibutuhkan sebuah teknologi peringkasan teks otomatis untuk menyederhanakan sebuah informasi evaluasi pembelajaran dosen yang panjang menjadi sebuah informasi yang sederhana tanpa menghilangkan inti dari teks tersebut. Teknologi peringkasan teks otomatis atau *text summarization* merupakan solusi untuk mengatasi permasalahan tersebut. Peringkasan teks otomatis adalah proses menghasilkan teks yang berasal dari satu atau banyak teks yang berisi informasi penting. Peringkasan teks otomatis memiliki dua metode, yaitu metode ekstraktif dan abstraktif. Metode ekstraktif bekerja dengan cara melakukan perhitungan skor untuk memilih beberapa kalimat yang penting pada dokumen dan menghasilkan sebuah ringkasan tanpa merubah isi daripada setiap kalimat tersebut. Sedangkan metode abstraktif yaitu dengan membuat kalimat baru yang tidak terdapat pada dokumen asli seperti yang manusia lakukan.

Pada tahun 2018 Ikhwan Nizwar Akhmad, dkk dari Universitas Sebelas Maret pernah membuat sebuah sistem peringkasan multidokumen otomatis dengan menggunakan *Log-Likelihood Ratio (LLR)* dan *Maximal Marginal Relevance (MMR)* untuk artikel bahasa Indonesia. Data yang digunakan dalam penelitian tersebut adalah artikel penyakit berbahasa Indonesia dari DokterSehat/AloDokter. Hasil penelitian ini menunjukkan bahwa *topic signature* (dan akurasi) sangat mempengaruhi hasil peringkasan otomatis dengan metode yang digunakan. Namun pada penelitian ini metode peringkasan ekstraktif memiliki kelemahan dalam pembentukan ringkasan yang koheren. Setiap kalimat belum membentuk hubungan antara satu sama lain[1].

Peringkasan abstraktif dipilih berdasarkan beberapa faktor diantaranya dataset yang digunakan dalam penelitian ini adalah evaluasi pembelajaran dosen di prodi Pendidikan Teknik Informatika Universitas Pendidikan Ganesha, dimana dataset tersebut ditulis oleh mahasiswa yang tata bahasa penulisannya tidak terjamin sesuai ejaan yang disempurnakan seperti pada artikel resmi. Oleh karena itu peringkasan otomatis dengan metode abstraktif dinilai cocok digunakan pada penelitian ini.

Pada tahun 2019 Rike Adelia, dkk dari Universitas Telkom pernah membuat sebuah *Indonesian Abstractive Text Summarization Using Bidirectional Gated Recurrent Unit*. Penelitian tersebut dilakukan dengan menggunakan metode Abstraktif. Dataset yang digunakan dalam penelitian ini adalah 500 dokumen jurnal berbahasa Indonesia dari berbagai sumber. Hasil yang didapat dari penelitian tersebut menunjukkan model mampu mempelajari dan memahami kata-kata yang terdapat pada dataset dan dapat menghasilkan ringkasan dengan kata-kata inti dari teks tersebut. Namun kelemahan pada penelitian ini menggunakan *Bidirectional Gated Recurrent Unit* yang mengakibatkan memiliki struktur tata bahasa yang buruk. Skor evaluasi dua skenario tidak lebih tinggi daripada skor model dalam bahasa Inggris. Hal ini disebabkan oleh ukuran dari teks dan faktor tata bahasa[2].

Pada tahun 2019 Nurrohmat, dkk pernah membuat sebuah sistem *Sentiment Analysis of Novel Review Using Long Short-Term Memory Method*. Dataset yang digunakan pada penelitian ini adalah data review novel berbahasa Indonesia yang diambil dari situs *goodreads.com*. Metode yang digunakan pada penelitian ini adalah metode *Long Short-Term Memory* dan *Naïve Bayes*. Penelitian ini membandingkan metode LSTM dengan metode Naïve Bayes berdasarkan perhitungan dari nilai *accuracy*, *precision*, *recall*, *f-measure*. Pada penelitian ini peneliti menyebutkan bahwa metode *Long Short-Term Memory* memiliki hasil akurasi yang lebih baik dibandingkan dengan metode *Naïve Bayes*[3].

Pada tahun 2020 Lionovan, dkk pernah membuat sebuah sistem Klasifikasi Topik dan Analisa Sentimen Terhadap Kuesioner Umpan Balik Universitas Menggunakan Metode *Long Short-Term Memory*. Dataset yang digunakan pada penelitian ini adalah kuesioner umpan balik di Universitas Kristen Petra. Metode yang digunakan pada penelitian ini adalah metode *Long Short-Term Memory*. Penelitian ini mengembangkan sebuah sistem pengklasifikasian topik dan analisa sentimen menggunakan *Word2vec* dan *Long Short-Term Memory*. Pada penelitian ini peneliti menyebutkan bahwa jumlah dan variasi komentar dapat mempengaruhi nilai rata-rata akurasi pada model LSTM. Selain itu juga penulis menyebutkan bahwa metode stemming saat melakukan proses preprocessing dapat meningkatkan nilai rata-rata akurasi pada model LSTM[4].

Pada tahun 2020 Yuliska, dkk pernah membuat sebuah Literatur Review Terhadap Metode, Aplikasi Dan Dataset Peringkasan Dokumen Teks Otomatis Untuk Teks Berbahasa Indonesia. Pada penelitian ini peneliti menyebutkan bahwa peringkasan dokumen teks secara otomatis didominasi oleh teknik yang bersifat ekstraktif. Peringkasan dokumen teks berbahasa Indonesia juga didominasi oleh metode-metode *unsupervised*, sementara metode-metode *supervised* seperti *machine learning* dan *deep learning* masih sangat jarang ditemukan[5].

Pada tahun 2019 Alpiani, dkk pernah membuat sebuah sistem *Pointer Generator Dan Coverage Weighting Untuk Memperbaiki Peringkasan Abstraktif*. Dataset yang digunakan pada penelitian ini adalah *CNN/Daily Mail* berbahasa Inggris. Metode yang digunakan pada penelitian ini adalah metode abstraktif. Pada penelitian ini peneliti menyebutkan bahwa secara keseluruhan model yang diusulkan menghasilkan sejumlah ringkasan yang lumayan mirip dengan ringkasan yang dibuat para ahli. Tetapi pada penelitian ini terdapat kekurangan, yaitu hasil ringkasan belum cukup signifikan jika dilihat dari sudut pandang tata bahasa. Hasil ringkasan dari model usulan ini bisa ditingkatkan dengan memperhitungkan informasi lain pada proses *training*[6].

Pada tahun 2019 Ivanedra, dkk pernah membuat sebuah sistem Implementasi Metode *Recurrent Neural Network Pada Text Summarization* Dengan Teknik Abstraktif. Dataset yang

digunakan pada penelitian ini adalah 4515 artikel atau berita berbahasa Inggris dari Hindu, Indian Times, dan Guardian. Metode yang digunakan pada penelitian ini adalah metode abstraktif. Pada penelitian ini peneliti menyebutkan bahwa perbedaan jumlah dataset yang digunakan sebagai bahan *training* sangat berpengaruh, karena pembendaharaan kata yang lebih banyak tentunya menjadikan program lebih mengerti dan lebih akurat dalam menciptakan ringkasan. Tetapi pada penelitian ini terdapat kekurangan, yaitu beberapa hasil ringkasan masih belum sesuai dengan konteks[7].

Pada tahun 2018 Yoko, dkk pernah membuat sebuah sistem Peringkasan Otomatis Abstraktif Dengan Menggunakan *Recurrent Neural Network*. Dataset yang digunakan pada penelitian ini adalah artikel atau berita berbahasa Indonesia dari Kompas/Detik. Metode yang digunakan pada penelitian ini adalah metode abstraktif. Pada penelitian ini peneliti menyebutkan bahwa ringkasan yang baik jika topik dari berita yang diringkas mirip dengan topik berita yang berada di dalam data latih walaupun kata-kata pada berita berbeda dengan data latih. Tetapi pada penelitian ini terdapat kekurangan, yaitu kesalahan yang dibuat sistem berupa ringkasan dengan kata-kata yang berhubungan subjek atau lokasi dari berita[8].

Selanjutnya pada tahun 2018 Kurniawan, dkk pernah membuat sebuah sistem *INDO SUM: A New Benchmark Dataset for Indonesian Text Summarization*. Dataset yang digunakan pada penelitian ini adalah 20 ribu artikel atau berita berbahasa Inggris dari berbagai sumber. Pada penelitian ini menghasilkan sebuah patokan dataset peringkasan ekstraktif yaitu INDOSUM. ROUGE merupakan standar evaluasi untuk teknologi peringkasan otomatis. Selain itu, peneliti juga menyebutkan bahwa perlu untuk fokus mengembangkan peringkasan dengan neural model yang terbaru untuk peringkasan abstraktif[9].

Berdasarkan beberapa hasil penelitian diatas, sentimen analisis dengan menggunakan metode LSTM cocok digunakan untuk melakukan pengklasifikasian evaluasi pembelajaran. Selanjutnya, peringkasan otomatis dengan menggunakan metode ekstraktif sudah banyak digunakan, sedangkan untuk peringkasan otomatis menggunakan metode abstraktif masih jarang sekali digunakan terutama yang menggunakan dataset berbahasa Indonesia. Peringkasan abstraktif dipilih berdasarkan beberapa faktor diantaranya dataset yang digunakan dalam penelitian ini adalah evaluasi pembelajaran dosen di prodi Pendidikan Teknik Informatika Universitas Pendidikan Ganesha, dimana dataset tersebut ditulis oleh mahasiswa yang tata bahasa penulisannya tidak terjamin sesuai ejaan yang disempurnakan seperti pada artikel resmi.

Pada penelitian peringkasan abstraktif sebelumnya seperti yang dilakukan oleh Rike Adelia menggunakan *Bidirectional Gated Recurrent Unit* menghasilkan struktur tata bahasa yang buruk. Skor evaluasi dua skenario tidak lebih tinggi daripada skor model dalam bahasa Inggris. Hal ini disebabkan oleh ukuran dari teks dan faktor tata bahasa. Pada penelitian ini

dikembangkan peringkasan abstraktif dengan menggunakan jaringan syaraf tiruan *Long Short Term Memory (LSTM)*. LSTM merupakan salah satu jenis dari *Recurrent Neural Network (RNN)* yang sudah dilakukan modifikasi dengan menambahkan *memory cell* agar dapat menyimpan informasi untuk jangka waktu yang lama[10].

Berdasarkan beberapa pertimbangan terhadap penelitian terdahulu diatas, maka pada penelitian ini peneliti mengembangkan sebuah sistem yang dapat melakukan pengklasifikasian sentimen dan peringkasan teks abstraktif untuk evaluasi pembelajaran dengan *algoritma LSTM(Long Short Term Memory)*.

II. KAJIAN TEORI

A. Sentiment Analysis

Sentiment Analysis adalah proses mempelajari data teks secara otomatis agar dapat menghasilkan informasi yang terdapat pada kalimat tersebut. Analisis sentimen biasanya digunakan untuk melihat pandangan suatu kalimat pada sebuah masalah, apakah berpendapat positif atau negatif. Contoh penggunaan analisis sentimen yaitu pada identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang[11].

Sentiment Analysis merupakan bagian dari ilmu Natural Language Processing (NLP) dan bergerak secara kontinum dimulai dari tahap klasifikasi teks, hingga tahap mereview polaritasnya[12].

B. Peringkasan Teks

Peringkasan teks merupakan salah satu cabang ilmu dari pemrosesan bahasa alamiah (*Natural Language Processing*). Peringkasan dokumen teks otomatis atau sering juga disebut sebagai *automatic text summarization* adalah sebuah cara untuk mengambil intisari informasi dari sebuah atau banyak dokumen teks[13]

Peringkasan teks dapat dilakukan dengan 2 cara, yaitu peringkasan teks secara ekstraktif (*Extractive summarization*) dan peringkasan teks secara abstraktif (*Abstractive summarization*). Peringkasan secara ekstraktif dilakukan dengan cara menghitung bobot setiap kata dan mengambil kata-kata yang memiliki bobot terbesar sebagai hasil ringkasan. Sedangkan peringkasan secara abstraktif dilakukan dengan cara membuat dan menyusun kalimat baru sebagai hasil ringkasan yang berisi intisari informasi dari dokumen yang diringkas seperti yang dilakukan oleh manusia.

C. Long Short Term Memory (LSTM)

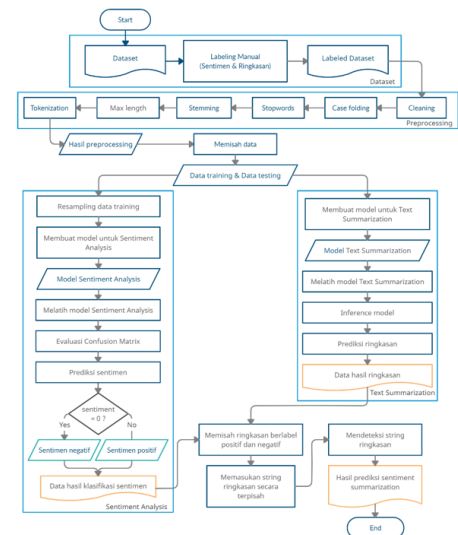
Pertama kali diperkenalkan oleh Hochreiter dan Schmidhuber pada tahun 1997. LSTM juga disebutkan sebagai jaringan syaraf tiruan dengan arsitektur yang mudah beradaptasi, sehingga bentuknya dapat disesuaikan tergantung pada aplikasinya[14].

Struktur *Long Short Term Memory (LSTM)* hampir sama dengan milik *Recurrent Neural Network (RNN)* yang memiliki bentuk rangkaian modul berulang dari jaringan syaraf tiruan. Perbedaannya adalah modul berulang pada RNN memiliki lapisan tunggal seperti lapisan tunggal Tanh. Sedangkan pada rangkaian LSTM memiliki empat lapisan yang saling berinteraksi[15].

Kegiatan sel state dikontrol oleh lapisan yang disebut dengan gerbang (*gates*). Gates terdiri dari lapisan jaringan syaraf sigmoid dan operasi multiplikasi yang searah, sehingga LSTM dapat menghapus atau menambahkan informasi kedalam sel. Inti dari *Long Short Term Memory* adalah sel state yang merupakan garis horizontal yang berjalan melewati bagian atas diagram. Sel state ini berjalan lurus ke bawah seluruh rantai dengan beberapa interaksi linear yang kecil. Hal ini yang menyebabkan informasi dapat mengalir dengan mudah tanpa adanya perubahan. Lapisan sigmoid (σ) mengeluarkan output berupa nilai antara 0 dan 1. Nilai 0 berarti tidak memperbolehkan informasi apapun masuk ke dalam sel, sedangkan nilai 1 berarti memperbolehkan informasi apapun masuk ke dalam sel[16].

III. METODOLOGI PENELITIAN

A. Flowchart Sistem



Gambar 1 Flowchart Sistem

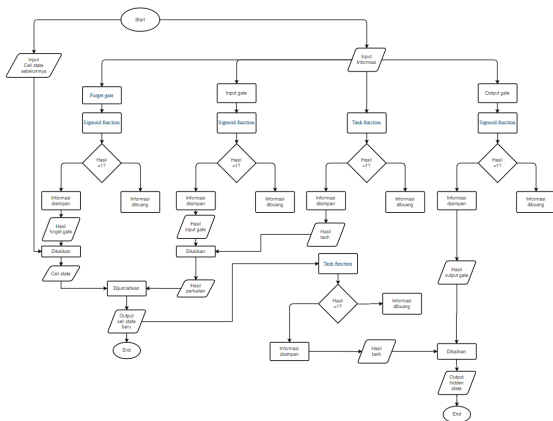
Pertama dataset yang telah memiliki label sentiment diberikan sebuah label tambahan yaitu ringkasan. Selanjutnya dataset tersebut dimasukan ke sistem dan dilakukan proses preprocessing melewati beberapa proses yaitu *cleaning*, *case folding*, *stopwords*, *stemming*, *max length*, dan *tokenization*. Lalu dataset tersebut dipisah menjadi *data training* dan *data testing* dengan rasio 8 berbanding 2.

Pada bagian *sentiment analysis*, dilakukan *resampling* untuk data *training* agar jumlah evaluasi positif dan evaluasi

negatif sama, tujuannya untuk meningkatkan akurasi hasil prediksi agar tidak memihak ke data dengan berlabel yang lebih banyak. Selanjutnya membuat model dan melatih model yang dilanjutkan dengan dilakukan evaluasi *confusion matrix*. Selanjutnya dilakukan prediksi untuk mencoba model. Jika hasil prediksi mengeluarkan nilai sama dengan nol maka sistem akan memprediksi sentimen pada evaluasi tersebut adalah negatif, sebaliknya jika tidak maka sentimennya adalah positif. Pada bagian *summarization*, dibuat sebuah model dan dilakukan *training*, lalu dilanjutkan dengan melakukan *inference model* hingga model dapat memprediksi ringkasan.

Evaluasi yang telah dilakukan klasifikasi sentimen dan diringkas, selanjutnya dipisah antara evaluasi berlabel positif dengan negatif. Selanjutnya memasukan string ringkasan secara terpisah, dan mendeteksi string ringkasan hingga string tersebut menjadi ringkas. Menghasilkan *output* dua buah ringkasan yaitu untuk label positif dan negatif.

B. Flowchart LSTM

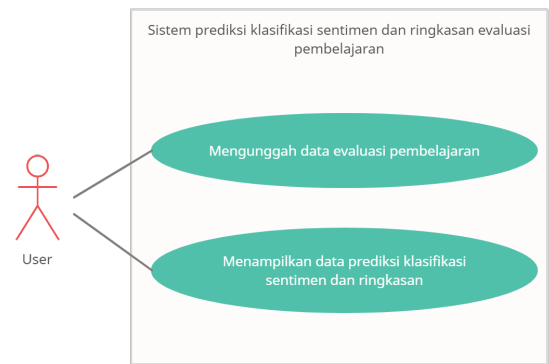


Gambar 2 Flowchart LSTM

Algoritma pada *LSTM layer* dijabarkan sebagai berikut, pertama nilai dari *cell state* sebelumnya akan masuk ke *cell* saat ini, jika *cell* saat ini merupakan *cell* pertama pada *LSTM layer* maka *cell state* sebelumnya bernilai 0. Disaat yang bersamaan informasi akan masuk dan melewati *forget gate* dimana pada *gate* ini akan dihitung dengan *sigmoid function* yang mana jika hasilnya 0 maka informasi tersebut dilupakan, sedangkan jika hasilnya 1 maka informasi tersebut akan diingat. Nilai dari *forget gate* akan dikalikan dengan nilai *cell state* sebelumnya menghasilkan nilai *cell state*. Informasi yang melewati *input gate* akan dihitung dengan *sigmoid function* lalu hasilnya dikalikan dengan hasil perhitungan informasi dengan *tanh function*. Selanjutnya nilai tersebut akan dikalikan dengan nilai hasil dari *input gate*, hasil perkalian tersebut selanjutnya dijumlahkan dengan nilai *cell state* menghasilkan nilai *cell state* baru, nilai tersebut diduplikasi menjadi dua, pertama akan diteruskan ke *cell* berikutnya, lalu kedua nilai *cell state* baru tersebut akan dihitung *tanh function* menghasilkan nilai *tanh*.

Informasi yang melewati *output gate* akan dihitung dengan *sigmoid function* lalu hasilnya akan dikalikan dengan nilai *tanh* menghasilkan nilai *hidden state* pada *cell* saat ini yang akan diteruskan ke *cell* berikutnya. Proses jalannya data tersebut terus diulangi hingga mencapai *cell* terakhir pada *LSTM layer*.

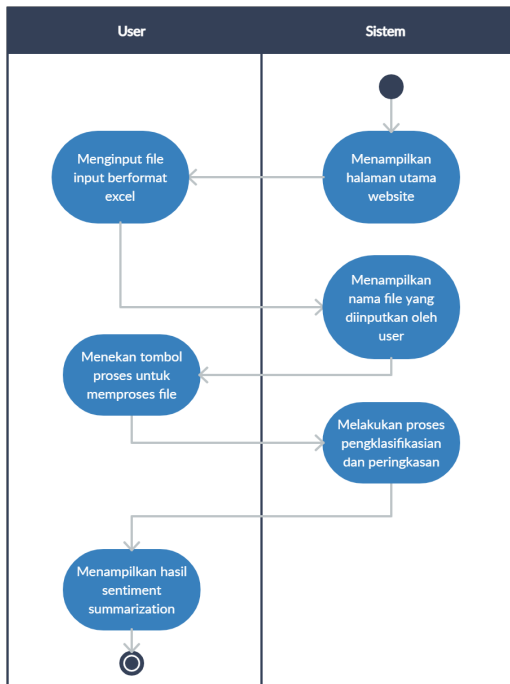
C. Use case diagram



Gambar 3 Use Case diagram

Use Case diagram pada sistem prediksi klasifikasi sentiment dan ringkasan evaluasi pembelajaran dimana user dapat menunggah data evaluasi pembelajaran ke sistem dan juga dapat mengunduh data prediksi klasifikasi sentiment dan dapat melihat data prediksi klasifikasi sentiment dan ringkasan yang ditampilkan oleh sistem.

D. Activity diagram



Gambar 4 Activity diagram

Activity diagram pada sistem prediksi klasifikasi sentiment dan ringkasan evaluasi pembelajaran dimana sistem akan menampilkan halaman utama website, lalu user menginputkan evaluasi dengan file berformat excel, selanjutnya sistem akan menampilkan nama file yang diinputkan oleh user, lalu user menekan tombol proses untuk memproses file, selanjutnya sistem akan melakukan proses pengklasifikasian dan peringkasan evaluasi, dan terakhir user akan melihat hasil sentiment summarization evaluasi pembelajaran yang ditampilkan oleh sistem.

E. Antarmuka

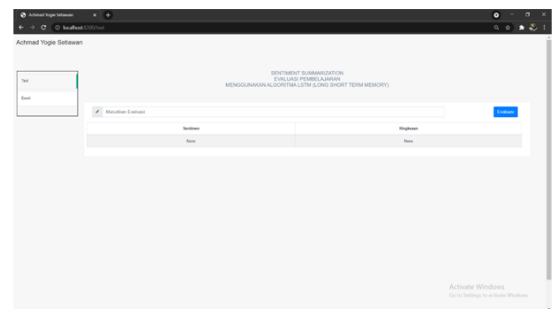


Gambar 5 Antarmuka

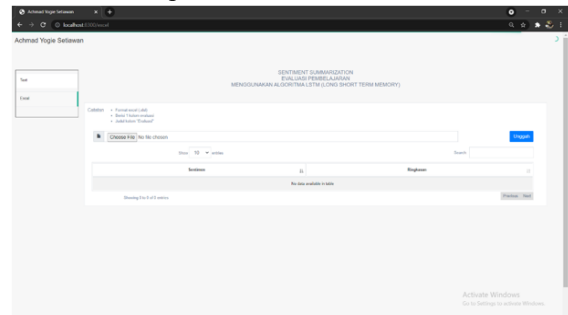
Antarmuka pada sistem prediksi klasifikasi sentiment dan ringkasan evaluasi pembelajaran terdapat tombol unggah file beserta catatan yang berisi ketentuan bentuk data agar dapat diproses oleh sistem. Jika pengguna telah mengunggah datanya yang berformat file Excel ke sistem, selanjutnya yaitu memilih tombol Proses agar sistem mulai memproses data yang diunggah pengguna. Jika data telah selesai diproses oleh sistem, maka tampilan antarmuka akan berubah dengan sistem menampilkan hasil sentiment summarization yang telah diprediksi.

IV. HASIL DAN PEMBAHASAN

A. Implementasi Antarmuka sistem



Gambar 6 Implementasi Antarmuka Sistem teks



Gambar 7 Implementasi Antarmuka Sistem excel

B. Pengujian Confusion Matrix

TABLE CONFUSION MATRIX

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	1080 (True Negatif)	105 (False Positif)
Aktual Positif	201 (False Negatif)	1880 (True Positif)

Setelah selesai dilakukan testing menggunakan data testing, dari 3266 evaluasi maka didapat bahwa terdapat 1080 evaluasi yang diprediksi bersentimen negatif dan cocok dengan label aktualnya. Lalu terdapat 201 evaluasi yang diprediksi bersentimen negatif dan tidak cocok dengan label aktualnya. Selanjutnya terdapat 105 evaluasi yang diprediksi bersentimen positif dan tidak cocok dengan label aktualnya. Terakhir terdapat 1880 evaluasi yang diprediksi bersentimen positif dan cocok dengan label aktualnya. Dari data tabel diatas, maka



KARMAPATI

dapat dicari nilai precision, recall, accuracy, dan f-measure sebagai berikut

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{1889}{1889 + 120} = 0,940$$

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{1889}{1889 + 192} = 0,907$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{1889 + 1065}{1889 + 1065 + 120 + 192} = 0,904$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall}$$

$$F - measure = \frac{2 \times 0,940 \times 0,907}{0,940 + 0,907} = 0,923$$

Confusion matrix yang dihasilkan dari library metrics dapat dilihat pada gambar berikut

	precision	recall	f1-score
Sentimen Negatif	0.847255	0.898734	0.872236
Sentimen Positif	0.940269	0.907737	0.923716
Accuracy	0.904470	0.904470	0.904470

Gambar 8 Classification Report

Accuracy merupakan ukuran seberapa akurat model dapat mengklasifikasikan dengan benar. Accuracy berisi rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Dengan kata lain, accuracy merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya).

Precision merupakan ukuran tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Precision berisi rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dari semua kelas positif yang telah di prediksi dengan benar, berapa banyak data yang benar-benar positif.

Recall merupakan ukuran keberhasilan model dalam menemukan kembali sebuah informasi. Recall berisi rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

C. Pengujian ROGUE

Metode untuk melakukan evaluasi hasil ringkasan teks secara otomatis yang paling sering digunakan yaitu ROUGE (Recall-Oriented Understudy for Gisting Evaluation). ROUGE

menghitung jumlah n-gram kata yang cocok antara ringkasan sistem dengan ringkasan referensi.

Pada penelitian ini proses uji coba ROGUE dilakukan dengan memberikan 20 evaluasi yang terdiri dari 10 evaluasi bersentimen positif dan 10 evaluasi bersentimen negatif yang selanjutnya evaluasi tersebut masing-masing diringkas oleh ahli bahasa. Selanjutnya ringkasan yang dibuat oleh ahli bahasa tersebut dibandingkan dengan ringkasan yang dihasilkan oleh sistem

TABLE EVALUASI POSITIF

No.	Evaluasi
1	Dalam mengajar di kelas, bapak sudah mengajar dengan sangat baik, yaitu dengan cara menjelaskan ataupun mencontohkan langsung di depan kelas sehingga mudah di ikuti dan di pahami oleh mahasiswa selain itu dalam mengajar sangat disiplin tetapi enjoy.
2	Ketika mengajar, Pak Wiguna sangat sistematis ketika memaparkan materi, khususnya dalam materi OOP. Sehingga kami paham dengan materi yang diberikan meskipun kadang dibuat bingung dengan pertanyaan yang sering beliau berikan. Saran saya kepada beliau adalah lebih bersabar lagi ketika menghadapi mahasiswa yang kurang disiplin ketika berada di kelas
3	Bapak sudah mengajarkan materi dengan baik, dan juga mengajari kita disiplin. semoga kedepannya lebih ditingkatkan lagi
4	Sudah bagus cara mengajarnya, mungkin perlu ditingkatkan lagi menjadi lebih seru dan kreatif sehingga kami semakin semangat untuk belajar, terima kasih
5	Saran:Selalu jadi dosen yang menyenangkan di kelas kami pak Kritik:Menurut saya bapak terlalu cepat dalam menerangkan sesuatu dimohonkan agar bapak lebih santai sehingga para mahasiswa memahaminya
6	Dalam proses pembelajaran sudah sangat baik, dan untuk sistem uas tergolong banyak jadi dimohon untuk pengurangannya. terimakasih
7	Cara mengajar dikelas kami bapak sudah cukup baik dan mudah dipahami oleh mahasiswa. Bapak orangnya friendly, memberikan kami motivasi. Saran untuk kedepannya bisa lebih dekat lagi dengan mahasiswa. Semangat dosen panutanku????
8	Cara mengajar sudah cukup baik dan menggunakan metode yg gampang dimengerti oleh mahasiswa, kedepannya hanya perlu ditingkatkan lagi agar semua mahasiswa merata bisa memahaminya?????????
9	Sudah cukup baik dalam mengajar dan memberi materi. Sehingga bisa dimengerti dan diikuti. Diharapkan untuk bisa lebih baik lagi kedepannya. Terimakasih
10	Cara mengajar bapak selama ini sudah sangat baik dan Selama saya diajarkan oleh bapak saya mendapatkan ilmu baru yang bermanfaat. juga dalam pemberian materi sangat menyenangkan dan tidak membuat saya bosan. Semoga kedepannya bisa lebih baik lagi dalam pengajarannya. Terimakasih

TABLE EVALUASI NEGATIF

No.	Evaluasi
-----	----------



1	Saran saya agar lebih ditingkatkan lagi dalam membawakan materi yang lebih kompleks sehingga mahasiswa bisa memahami dengan cepat, kritik saya supaya materi yang diajarkan dibawakan lebih menarik agar mahasiswa semangat dalam mengikuti pembelajaran
2	Cara mengajarnya sudah tersusun secara rapih, namun terkadang terlalu cepat. Sarannya sebaiknya memastikan tidak ada mahasiswa yang tertinggal dalam 1 tahapan materi.
3	Saran saya adalah mungkin untuk wifi di lab dasar agar lebih ditingkatkan agar saat mengerjakan uts maupun uas oracle tidak terdapat kendala mengenai jaringan internet. Kritik saya adalah menurut saya tugas project yang diberikan untuk uas terlalu banyak
4	Memberikan lebih banyak lagi contoh-contoh latihan tentang materi yang sedang di berikan, agar mahasiswa dapat lebih memahami materi yang di sedang di berikan
5	Saran saya dalam melaksanakan perkuliaha agar menjelaskan materi dengan detail dan menyarankan buku untuk pembelajaran
6	Saran : pada saat proses pembelajaran sedang berlangsung interaksi antara peserta didik harus sering dilakukan guna mempermudah peserta didik dalam memahami materi. Dan jika peserta didik belum paham akan materi tersebut agar dilakukan tanya jawab agar peserta didik dengan baik memahami materi yang diajarkan. Kritik : pada saat proses pembelajaran yang berlangsung sudah baik. Akan tetapi saat proses pembelajaran berlangsung peserta didik lebih di kontrol seperti melakukan tanya jawab, mengajukan pertanyaan apabila ada yg belum dimengerti, agar peserta didik bisa memahami materi dengan baik.
7	Cara mengajarnya sudah tersusun secara rapih, namun terkadang terlalu cepat. Sarannya sebaiknya memastikan tidak ada mahasiswa yang tertinggal dalam 1 tahapan materi.
8	Kritik saya pada mata kuliah ini adalah dosen masih sedikit meluangkan waktunya untuk memberikan kesempatan kepada mahasiswa untuk berdiskusi. Kemudian terdapat batasan antara hubungan dosen dengan mahasiswa. Saran saya sebelum memulai perkuliahan dosen harus nya menyusun rancangan perkuliahan terlebih dahulu dimana dosen seharusnya memberikan waktu lebih banyak kepada mahasiswa untuk mendiskusikan tgs.
9	Om Swastyastu, Saya ingin memberi sedikit saran kepada bapak pada saat dikelas terkadang terlalu cepat menjelaskan dan saya kadang sulit untuk mengikutinya. Jadi untuk kedepannya mungkin bisa lebih dipelankan lagi untuk pengajaran dikelas. Sisanya sudah cukup baik menurut saya, cukup mudah dimengerti namun terkadang juga membuat bingung dikelas karena terlalu cepat menurut saya.
10	Saran dan kritik saya terhadap dosen saya yang mengajar mata kuliah Pemrograman Berorientasi Objek yaitu bapak I Gede Mahendra Darmawiguna, S.Kom.,M.Sc. adalah yang pertama saran saya adalah bapak supaya lebih tegas dalam mengajar ke mahasiswanya dan kritik saya adalah ketika bapak mengajarkan materi tentang sesuatu diharapkan bapak mengajarkannya secara pelan-pelan dan jangan cepat-cepat agar supaya mudah dimengerti oleh para mahasiswanya

TABLE UJI ROGUE EVALUASI POSITIF

Sistem	Referensi (Manusia)
pembelajaran yang disampaikan materi diajarkan banyak contoh dirasa sebaiknya lebih tegas dan agar mahasiswa aktif dengan baik	Cara mengajar sudah baik, mudah dipahami, dan menyenangkan. Selain itu, bapak disiplin dan friendly sehingga mahasiswa termotivasi untuk belajar

$$ROUGE - 1 = \frac{3}{18} = 0,16$$

Pada evaluasi negatif, hanya terdapat 3 dari 18 kata yang sama antara prediksi ringkasan yang dibuat oleh sistem dengan ringkasan yang dibuat oleh ahli bahasa. Kata-kata yang sama tersebut yaitu diantaranya “baik”, “dan”, dan “mahasiswa”.

TABLE UJI ROGUE EVALUASI POSITIF

Sistem	Referensi (Manusia)
sebaiknya sudah harus tegas lebih terutama proses pembelajaran metode mengajar lagi dalam kritik dan slide banyak saran kedisiplinan	Penjelasan materi terlalu cepat, kurang diskusi, dan terlalu banyak memberikan tugas

$$ROUGE - 1 = \frac{2}{10} = 0,2$$

Pada evaluasi positif, hanya terdapat 2 dari 10 kata yang sama antara prediksi ringkasan yang dibuat oleh sistem dengan ringkasan yang dibuat oleh ahli bahasa. Kata-kata yang sama tersebut yaitu diantaranya “banyak” dan “dan”.

V. KESIMPULAN

Pada pengujian confusion matrix sistem mendapatkan nilai accuracy sebesar 0,902 dan nilai f-measure sebesar 0,921. Nilai tersebut berarti sistem sudah dapat menghasilkan prediksi sentimen yang sesuai dengan data latih.

Pada pengujian Recall-Oriented Understudy for Gisting Evaluation (ROGUE) evaluasi positif mendapatkan nilai 0,16 dan pada evaluasi negatif mendapatkan nilai 0,2. Nilai tersebut berarti sistem belum dapat menghasilkan prediksi ringkasan yang sesuai secara maksimal dengan ringkasan milik ahli.

Nilai ROGUE yang kecil didapat karena teknik penulisan ringkasan yang berbeda antara yang peneliti pada dataset yang dilatihkan ke model dengan ringkasan yang dibuat oleh ahli bahasa. Perbedaan teknik meringkas tersebut membuat jumlah kata yang sama sedikit mengakibatkan nilai ROGUE menjadi kecil. Selain itu juga kecilnya nilai ROGUE diakibatkan karena jumlah *epochs* saat *training model summarization* masih kecil yaitu hanya 50 dengan hasil nilai *accuracy* hanya 0.8993 dan *loss* 0.3404. Besarnya dataset membuat proses *training* memakan waktu yang lama. Pada penelitian ini peneliti dengan *resource* dan *hardware* yang terbatas hanya mampu melakukan *training* dengan 50 *epochs*. Jika dilakukan dengan *resource* dan *hardware* yang lebih memadai dapat melakukan *training* dengan *epochs* yang lebih banyak sehingga menghasilkan nilai *accuracy* dan *loss* yang lebih bagus sehingga menghasilkan prediksi peringkasan yang lebih akurat sesuai dengan data *training*.

Tokenizer yang dikembangkan belum dapat menyimpan token hasil proses *training*. Jika sistem dibuka kembali dengan mengunggah model yang tersimpan, maka sistem akan



KARMAPATI

Kampus Arsitektur, Riset, dan Inovasi

kesulitan dalam menentukan token pada setiap kata. Permasalahan tersebut berdampak pada proses prediksi ringkasan yang dilakukan oleh *model summarization*, dimana *input user* harus melewati *tokenizer* terlebih dahulu sebelum dapat diproses. Akibatnya, *token* pada setiap kata menjadi acak dan tidak sesuai dengan yang telah dipelajari model sebelumnya.

Dalam pengembangan selanjutnya, diharapkan dapat menguji dengan jumlah dataset yang lebih banyak, karena semakin banyak data yang dipakai maka sistem akan semakin baik dalam mempelajari pola data yang digunakan sehingga prediksi yang dihasilkan oleh sistem akan semakin akurat. Lalu saat pemberian label ringkasan untuk data latih dapat dilakukan oleh orang atau teknik yang sama dengan saat uji coba ROGUE agar gaya penulisannya lebih mirip. Lalu juga dapat melakukan perbandingan kinerja model dengan parameter yang berbeda guna mencari model terbaik yang dapat menghasilkan prediksi terakurat. Lalu juga dapat menganalisis algoritma lain yang sekiranya dapat lebih baik dalam melakukan prediksi klasifikasi sentimen dan ringkasan evaluasi pembelajaran. Selain itu juga dapat mengembangkan *tokenizer* yang dapat menyimpan *token* setelah dilakukan proses *training* agar nilai *token* tidak berubah-ubah dan dapat digunakan kembali tanpa harus mengulangi proses *training*. Lalu juga dapat membandingkan hasil prediksi sistem dengan menggunakan *framework deep learning* yang lain seperti PyTorch, MXNet, ataupun ONNX.

REFERENCES

[1] Akhmad, I. N. dkk. 2018. "Peringkasan Multidokumen Otomatis dengan Menggunakan Log-Likelihood Ratio (LLR) dan Maximal Marginal Relevance (MMR) untuk Artikel Bahasa Indonesia". *Jurnal Linguistik Komputasional*. Vol.1, No.1. Terdapat pada <http://inacl.id/journal/index.php/jlk/article/view/6/7>. Diakses pada 15 Desember 2020.

[2] R. Adelia, S. Suyanto and U.N. Wisesty, "Indonesian Abstractive Text Summarization Using Bidirectional Gated Recurrent Unit," *International Conference on Computer Science and Computational Intelligence*, 2019

[3] Nurrohmat, M. A., dkk. "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," in *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 2019.

[4] Lionovan, D. A., dkk. 2020. "Klasifikasi Topik dan Analisa Sentimen Terhadap Kuesioner Umpan Balik Universitas Menggunakan Metode Long Short-Term Memory". *Jurnal Infra*. Vol.9, No.2. Terdapat pada

<http://publication.petra.ac.id/index.php/teknik-informatika/article/view/10512/9355>. Diakses pada 19 November 2020.

[5] Y. Yuliska and T. Sakai, "A Comparative Study of Deep Learning Approaches for QueryFocused Extractive Multi-Document Summarization," in *International Conference on Information and Computer Technologies*, 2019, p.p. 153-157.

[6] Alpiani, A. S. dkk. 2019. "Pointer Generator dan Coverage Weighting untuk Memperbaiki Peringkasan Abstraktif". *Ind. Journal on Computing*. Vol.4, No.2. Terdapat pada <https://socj.telkomuniversity.ac.id/ojs/index.php/indoic/article/view/341/148>. Diakses pada 15 Desember 2020.

[7] Ivanedra, K. dkk. 2019. "Implementasi Metode Recurrent Neural Network Pada Text". *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*. Vol.6, No.4. Terdapat pada <http://jtiik.ub.ac.id/index.php/jtiik/article/view/1067>. Diakses pada 15 Desember 2020.

[8] Yoko, K. dkk. 2018. "Sistem Peringkas Otomatis Abstraktif Dengan Menggunakan Recurrent Neural Network". *Journal of Computer Science and Information Systems*. Vol.2, No.1. Terdapat pada <https://journal.untar.ac.id/index.php/computatio/article/view/1481>. Diakses pada 15 Desember 2020.

[9] Kurniawan, K., Louvan, S. 2018. INDO SUM: A New Benchmark Dataset for Indonesian Text Summarization. *Proceedings of 2018 International Conference on Asian Language Processing (IALP)*, Bandung: 15-17 November 2018.

[10] Zaman, L., Sumpeno, S., Hariadi, M., "Analisis Kinerja LSTM dan GRU sebagai Model Generatif untuk Tari Remo," in *JNTETI*, 2018.

[11] Rozi, I. F., Pramono, S. H, Dahlan, E. A., "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," in *Jurnal EECCIS*, 2012.

[12] Kouloumpis, E., Wilson, T., Moore, J., "Twitter Sentiment Analysis: The Good the Bad and the OMG!," in *ICWSM*, 2011.

[13] Yuliska. dkk. 2020. "Literatur Review Terhadap Metode, Aplikasi dan Dataset Peringkasan Dokumen Teks Otomatis untuk Teks Berbahasa Indonesia". *IT Journal Research and Development (ITJRD)*. Vol.5, No.1. Terdapat pada <https://journal.uir.ac.id/index.php/ITJRD/article/view/4688/2567>. Diakses pada 17 November 2020.

[14] Wiranda, L., Sadikin, M, "PENERAPAN LONG SHORT TERM MEMORY PADA DATA TIME SERIES UNTUK MEMPREDIKSI PENJUALAN PRODUK PT. METISKA FARMA," in *Jurnal Nasional Pendidikan Teknik Informatika*, 2019.

[15] Sanjeevi, M. 2018. "Chapter 10.1: DeepNLP — LSTM (Long Short Term Memory) Networks with Math.". <https://medium.com/deep-math-machine-learning-ai/chapter-10-1-deepnlp-lstm-long-short-term-memory-networks-with-math-21477f8e4235>. Diakses pada 22 Januari 2021 jam 09.05 WITA.

[16] Colah. 2015. "Understanding LSTM Networks". <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Diakses pada 22 Januari 2021 jam 12.32 WITA.