

Klasifikasi Sentimen Opini Publik Pada Instagram Pemerintah Kabupaten Bojonegoro Menggunakan LSTM

Titis Arwindarti¹, Esther Irawati Setiawan^{2*}, Syaiful Imron³

^{1,2,3} Program Studi Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya, Jawa Timur
Email: ¹ titis_a20@mhs.istts.ac.id, ^{2*} esther@stts.edu, ³ imron_i20@mhs.istts.ac.id

(Naskah masuk: 3 Okt 2023, direvisi: 27 Okt 2023, diterima: 1 Nov 2023)

Abstrak

Media sosial banyak membantu masyarakat dalam mendapatkan informasi terbaru terkait peristiwa atau kejadian dilingkungan sekitar maupun lebih luas. Masyarakat dapat menyampaikan pendapat mereka melalui tulisan dan dapat mengekspresikannya melalui fitur *emoticon* pada *platform* media sosial. Pemerintah Kabupaten Bojonegoro menggunakan *platform* Instagram sebagai salah satu sarana dalam menyampaikan informasi kepada masyarakat. Selaku pembuat kebijakan pelayanan publik membutuhkan *feedback* dari masyarakat agar kebijakan yang dibuat bisa tepat sasaran dan bermanfaat bagi masyarakat. Sentimen opini publik merupakan aspek penting dalam memahami respon masyarakat terhadap layanan masyarakat, program dan kebijakan yang dibuat. Peneliti mengumpulkan dan mengolah data yang diperoleh dari proses *scrapping* akun resmi Instagram Pemerintah Kabupaten Bojonegoro sebanyak 4.637 *dataset* yang selanjutnya dilakukan pelabelan data. Penelitian ini menggunakan *word embedding* Word2Vec untuk mengubah teks menjadi representasi vektor dan *Long Short-Term Memory* (LSTM) untuk melakukan klasifikasi. Dengan menggunakan *confusion matrix* menunjukkan bahwa model LSTM yang dibuat hasilnya mencapai akurasi 84,16%. Hasil analisa tersebut dapat memberikan kontribusi positif dan dapat menjadi bahan pertimbangan Pemerintah Kabupaten Bojonegoro dalam upaya meningkatkan layanan masyarakat, program dan kebijakan yang dibuat.

Kata Kunci: Analisa Sentimen, Opini Publik, Instagram, LSTM, Word2vec.

Classification of Public Opinion Sentiment on Instagram for Bojonegoro Regency Government Using LSTM

Abstract

Social media greatly helps society in obtaining the latest information regarding local or broader events and incidents. People can express their opinions through written posts and can convey their emotions using emoticon features on social media platforms. The Bojonegoro Regency Government uses Instagram as one of the means to communicate information to the public. As policymakers in public service, they require feedback from the community to ensure that the policies they create are targeted and beneficial to the public. Public sentiment is a crucial aspect of understanding the community's response to public services, programs, and policies. Researchers collected and processed data obtained from scraping the official Instagram accounts of the Bojonegoro Regency Government, resulting in 4,637 datasets that were subsequently labeled. This study used Word2Vec word embedding to transform text into vector representations and Long Short-Term Memory (LSTM) for classification. Using a confusion matrix, it is evident that the LSTM model achieved an accuracy of 84.16%. The results of this analysis can provide a positive contribution and serve as valuable input for the Bojonegoro Regency Government in their efforts to enhance public services, programs, and policies.

Keywords: *Sentiment Analysis, Public Opinion, Instagram, LSTM, Word2vec.*

I. PENDAHULUAN

Analisis pendapat pengguna media sosial selalu menjadi perhatian besar dalam mendukung pengambilan keputusan diberbagai bidang [1]. Saat ini, analisis sentimen melibatkan media apasaja yang digunakan masyarakat dalam mengekspresikan pendapat mereka seperti tulisan, ucapan, gerakan, ekspresi wajah, tindakan dan lain sebagainya [2]. Emoji yang terdiri dari berbagai simbol mulai dari ekspresi wajah kartun hingga *figure* seperti bendera dan olahraga banyak digunakan dalam komunikasi sehari-hari untuk mengungkapkan perasaan Masyarakat [3]. Media sosial telah membantu masyarakat untuk mendapatkan informasi terbaru terkait peristiwa atau kejadian dilingkungan sekitar maupun yang lebih luas. Semakin maraknya penggunaan teknologi *smartphone* menjadikan masyarakat dengan mudah menyampaikan opini melalui media sosial ke akun Instansi Pemerintah [4]. Penilaian sentimen digunakan untuk meninjau sudut pandang tentang layanan yang diungkapkan dalam komentar [5]. Atas dasar hal tersebut, banyak pemerintah daerah menggunakan media sosial menjadi salah satu layanan *electronic government (E-Government)* sebagai sarana dalam menyampaikan informasi kepada Masyarakat [6]. Informasi yang disampaikan seringkali mendapat *feedback* dari masyarakat, banyaknya opini tersebut dapat menjadikan salah satu sumber informasi kembali kepada pemerintah selaku pelayanan publik dalam membuat kebijakan sesuai dengan kebutuhan masyarakat sehingga dapat menjadikan landasan dalam pembuatan keputusan yang lebih tepat sasaran [7].

Berdasarkan dataindonesia.id jumlah pengguna media sosial di Indonesia terus meningkat pada setiap tahunnya. Jumlah pengguna aktif media sosial di Indonesia sebanyak 191 juta orang pada bulan Januari tahun 2022. Jumlah tersebut meningkat 12,35% dari tahun sebelumnya. WhatsApp menjadi media sosial yang paling banyak digunakan masyarakat Indonesia dimana mencapai 88,7 % dan selanjutnya Instagram yang mencapai 84,8% [8].

Penelitian tentang analisis sentimen berdasarkan data teks dari media sosial banyak dilakukan, salah satunya menggunakan metode *machine learning Support Vector Machine (SVM)* yang melakukan analisis sentimen pada kolom komentar Instagram untuk mengetahui sentimen dari setiap komentar dengan objek *cyberbullying* (perundungan di internet) [9]. Selain itu terdapat penelitian yang membandingkan SVM dengan LSTM *Classifier* untuk menganalisis sentimen tokoh publik di twitter [10]. Pada umumnya algoritma berbasis mesin pembelajaran seperti *Neural Network*, *k-Nearest Neighbor*, SVM dan LSTM digunakan sebagai metode dalam klasifikasi teks [11]. Jika berdasarkan pendekatan ekstraksi fitur SVM, nilai akurasi tertinggi terdapat pada perhitungan dengan menggunakan ekstraksi fitur *Count Vector* dan TF-IDF N-Gram *level* dengan nilai akurasi 59% pada instagram. SVM belum bisa digunakan sebagai algoritma yang menangkap urutan karena kata yang dibolak-balik meskipun maknanya berbeda tetap diakui sama oleh mesin dengan klasifikasi SVM, diperlukan algoritma klasifikasi yang bisa menangkap urutan kata dengan makna

yang sesuai, misalnya dengan menggunakan algoritma *Highest Response Ratio Next (HRRN)* yang memperhatikan proses dengan rasio respon paling tinggi atau *Long Short-Term Memory (LSTM)* [12]. Penelitian ini menggunakan *word embedding Word2Vec* untuk mengubah teks menjadi representasi vektor dan *Long Short-Term Memory (LSTM)* untuk melakukan klasifikasi sentimen opini publik pada akun Instagram Pemerintah Kabupaten Bojonegoro. *Sentiment analysis* digunakan untuk menganalisa teks pada komentar yang dikemukakan masyarakat dalam akun tersebut. Penelitian ini menggunakan bahasa pemrograman python dalam memprediksi komentar yang berupa positif, netral dan negatif. Hasil analisa tersebut dapat memberikan kontribusi positif dan dapat menjadi bahan pertimbangan Pemerintah Kabupaten Bojonegoro dalam upaya meningkatkan layanan masyarakat, program dan kebijakan yang dibuat.

II. TINJAUAN PUSTAKA

Pada penelitian sebelumnya [13], penulis menganalisis sentimen kepuasan mahasiswa terhadap pusat layanan LaC (*Language Center*) Universitas Telkom di Instagram. Penulis melakukan klasifikasi berdasarkan kategori positif, negatif dan netral menggunakan metode *Reccurent Neural Network (RNN)* dengan pengukuran *Confusion Matrix*. Metodologi pada tahap pertama adalah pengumpulan data dari komentar Instagram, tahap kedua memberi label *dataset* yang telah dibuat secara manual, ketiga tahap *preprocessing*, tahap keempat pemodelan menggunakan LSTM dan kelima proses klasifikasi RNN. Berdasarkan hasil pengujian dengan metode tersebut disimpulkan hasil pengujian akurasi dengan parameter yang ditentukan mendapatkan akurasi tertinggi sebesar 79% dan akurasi terendah 61%. Sistem mendapatkan nilai rata-rata berdasarkan parameter presisi tertinggi 78%.

Pada penelitian [14], penulis menganalisa reaksi masyarakat terhadap wabah virus menggunakan *hashtag* di platform Instagram. Posting pada Instagram dianalisa untuk memunculkan kesamaan cara individu dalam memberikan reaksi terhadap krisis Covid-19. Dalam pengumpulan data penulis menggunakan sebuah *Python Package Index (PyPI) igranscrapper* dengan menggunakan tagar #wuhan, #corona, dan #viruscorona. Penulis mengambil data dari bulan Februari tahun 2016 hingga bulan Maret tahun 2020, disimpan dalam *file excel* dan data media video hanya diambil dari *frame* pertama. Tahap selanjutnya adalah pembersihan data kemudian dikategorikan menjadi lima bagian antara lain meme / humor, berita / netral, positif, negatif dan acak. Penulis menggunakan *Google Lens* dalam menerjemahkan konten media dan *Google Translate* dalam menerjemahkan teks. LSTM digunakan untuk melatih data teks dalam analisis sentimen berupa teks dan klasifikasi media menggunakan *CNN ResNet 50*. Hasil yang diperoleh dari eksperimen tersebut untuk sentimen tekstual menggunakan LSTM penulis mencapai akurasi 87,7% kemudian saat menerapkan *transfer learning* dan melatih model penulis mencapai akurasi 73,6%. Dalam melakukan analisis sentimen media, penulis mendapat akurasi 82,1% saat melatih data dan setelah disempurnakan

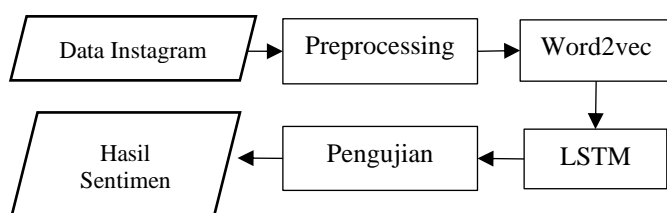
dengan CNN Resnet-50 data tersebut mencapai akurasi 79,2%.

Pada penelitian [15], penulis mengumpulkan data dari ulasan film IMDB. Penulis membagi data 80% sebagai data pelatihan dan 20% sebagai data pengujian untuk menentukan hasil model yang dipelajari. Tujuan dari penelitian ini adalah menentukan apakah *review* dari ulasan tersebut bersifat positif atau negatif dalam hal polaritas *review* yang diberikan. Penulis menggunakan beberapa teknik, antara lain *Naïve Bayes*, *Support Vector Machine (SVM)*, *Logistic Regresion* dan *Recurrent Neural Network (RNN)* untuk mengklasifikasi teks. Dari algoritma yang dilakukan dalam penelitian ini, algoritma RNN mencapai akurasi 87,70%, algoritma *Naïve Bayes* mencapai akurasi 75,1%, algoritma SVM mencapai akurasi 58,22% dan *Regresi Logistik* mencapai 75,18%.

Pada penelitian [16], penulis menganalisa sentimen opini publik pada Instagram dengan algoritma *convolutional neural network*. Dalam penelitian tersebut, penulis melakukan *training* menggunakan *Convolutional Neural Network (CNN)* untuk ekstraksi fitur yang terdapat pada teks. Setiap kata direpresentasikan dengan *word embedding* untuk dijadikan input pada CNN. Proses pada teks menggunakan SenticNet 5 dan dilakukan ekstraksi konsep untuk mendapatkan nilai *polarity* berdasarkan konsep yang terdapat pada kalimat tersebut. Kalimat tersebut diberi *padding* untuk mendapatkan jumlah yang sama pada setiap kalimat agar dapat dianalisa pada model yang dibuat. Untuk mengubah kalimat menjadi *vector*, penulis menggunakan *Global Vector (GloVe)*. Arsitektur CNN yang digunakan dalam penelitian ini menggunakan 1 *convolutional layer* dengan ukuran 2x2 dengan 100 *kernel*. Setiap *convolutional layer* diikuti dengan *max-pool layer* dengan ukuran 2x2, terdapat *fully connected layer* dengan 500 *neuron* dan diakhiri dengan *softmax layer* dengan 4 *neuron* untuk mewakili *class aspek* dan 3 *neuron* untuk mewakili *class sentiment*. Hasil dari penelitian menunjukkan 4 aspek ((Layanan, Kecepatan, *Coverage*, dan Promosi) dan 3 sentimen (Positif, Netral, dan Negatif) serta mendapatkan nilai rata-rata *precision* sebesar 95,80%, *recall* sebesar 88,12% dan *f1-score* sebesar 91,62%.

III. METODE PENELITIAN

Metode penelitian pada penelitian ini dilakukan dengan cara pengumpulan data pada akun Instagram Pemerintah Kabupaten Bojonegoro, *preprocessing*, *word Embedding*, pembuatan model menggunakan algoritma LSTM, pengujian dan *output* hasil sentimen. Gambar 1 merupakan desain arsitektur dalam penelitian ini.



Gambar 1. Desain Penelitian

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dengan cara *scrapping* menggunakan *Apify Instagram Scraper* yang disimpan kedalam format *excel* dan data yang diambil hanya berupa teks deskriptif dalam Bahasa Indonesia. Data yang diproses hanya data yang sesuai dengan topik yang dibahas pada *caption*. Setiap *caption* diberi label bidang dimana setiap bidang terdiri dari pengelompokan Instansi sesuai bidang yang berkepentingan. Opini diberi label sentimen menjadi positif, negatif dan netral.

B. Data Preprocessing

Pada penelitian ini menerapkan data *preprocessing* dalam melakukan pengolahan data, pada tahap ini merupakan proses yang berfungsi untuk membersihkan teks sebelum dilakukan pengolahan lebih lanjut. Data hasil *scrapping* yang diperoleh biasanya tidak berstruktur serta terdapat banyak *noise* seperti tanda baca, imbuhan, angka, karakter-karakter khusus dan lain sebagainya. Pada tahap ini, data akan dibersihkan sehingga tersisa bentuk dasar dari masing-masing kata saja untuk keperluan analisis lebih lanjut [17]. Terdapat beberapa tahapan yang dilakukan dalam fase *preprocessing* sehingga data yang digunakan dapat siap untuk diolah pada fase berikutnya. Beberapa tahapan diantaranya adalah sebagai berikut:

1) Case Folding

Case Folding bertujuan untuk membuat semua teks menjadi huruf kecil.

2) Removal URL

Removal URL bertujuan untuk menghapus *URL* atau alamat web yang terdapat pada komentar Instagram dimana hal tersebut tidak efektif dan tidak memiliki arti. Kemunculan *URL* ini disebabkan karena banyaknya pengguna mengarahkan membuka situs halaman web yang berkaitan dengan *caption* Instagram.

3) Remove Punctuation

Remove Punctuation bertujuan menghapus semua karakter selain alfabet misalnya simbol, spasi dan lain-lain.

4) Remove Hashtag

Remove Hashtag bertujuan menghapus tanda "#". *Hashtag* merupakan petunjuk kata yang sedang dibicarakan atau biasanya pengelompokan terhadap percakapan oleh sesama pengguna Instagram yang memiliki simbol "#".

5) Remove Mention

Remove Mention digunakan untuk menghapus simbol @ dimana simbol tersebut digunakan untuk menunjuk atau mengajak teman berkomunikasi langsung pada akun Instagram.

6) Convert Emoticon

Convert Emoticon digunakan untuk mengubah ekspresi yang diungkapkan menggunakan gambar (*emoji*) kedalam bentuk *string* yang dapat diartikan maknanya.

7) Convert Word

Convert Word digunakan untuk mengubah kata tidak baku menjadi kata baku yang sesuai dengan KBBI karena banyaknya penggunaan bahasa gaul.

8) *Remove Stopword*

Remove Stopword digunakan untuk menghilangkan kata yang sering keluar namun dianggap tidak penting seperti waktu, penghubung dan lain sebagainya.

9) *Stemming*

Stemming digunakan untuk mendapatkan kata dasar dari suatu kata. Hal ini dilakukan untuk menormalisasi kata.

10) *Tokenisasi*

Tokenisasi merupakan proses mengonversi teks dalam bentuk huruf menjadi urutan angka atau token sehingga dapat diproses oleh model LSTM.

C. *Word Embedding*

Setelah dilakukan *preprocessing*, tahap berikutnya adalah *word embedding*. *Word embedding* merupakan sebuah metode yang digunakan untuk membuat vector dengan mempresentasikan kata yang diperkirakan memiliki arti yang sama [18]. Ekstraksi fitur data teks menggunakan vektor Word2Vec merupakan proses menghitung rata-rata vektor dari setiap kata yang ada dalam sebuah kalimat [19] dan ditulis dengan rumus (1).

$$V_{sentence}(W) = \frac{1}{n} \sum_{i=1}^n v_{wi} \tag{1}$$

Keterangan :

w = kalimat

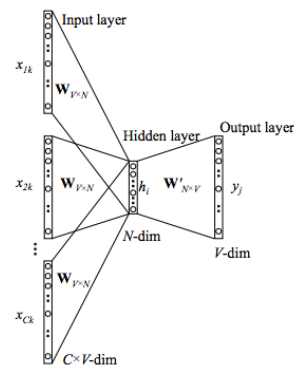
n = jumlah kata pada kalimat w

v_{wi} = vektor dari kata w_i

Word2Vec diperkenalkan oleh Mikolov dkk. pada tahun 2013. Model Word2Vec ini digunakan untuk memahami hubungan semantik antar kata dan merupakan salah satu aplikasi *unsupervised learning* menggunakan *neural network* yang terdiri dari sebuah *hidden layer* dan *fully connected layer*. Dimensi dari matriks bobot pada setiap *layer* adalah jumlah kata dalam korpus dikalikan dengan jumlah *hidden neuron* pada *hidden layer*-nya. Matriks bobot pada *hidden layer* dari model yang telah dilatih digunakan untuk mentransformasikan kata ke dalam vektor. Matriks bobot ini seperti *lookup table*, di mana setiap baris mewakili setiap kata dan kolom mewakili vektor dari kata tersebut. Semantik yang dipelajari dari kata tertentu dipengaruhi oleh kata-kata sekitarnya. Model ini mendemonstrasikan kemampuan untuk mempelajari pola linguistik sebagai hubungan linear antar vektor kata [20]. Terdapat dua algoritma Word2Vec yaitu *Continuous Bag of Word (CBOW)* dan *Skip-gram*.

1) *CBOW*

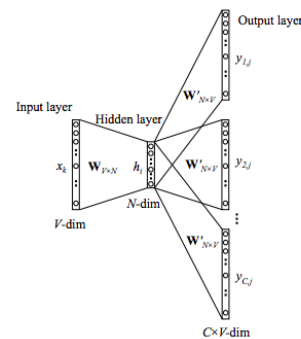
CBOW merupakan model yang menggunakan konteks untuk memprediksi target kata. CBOW merupakan model yang dipilih dalam penelitian ini karena memiliki waktu training lebih cepat dan memiliki akurasi yang sedikit lebih baik untuk *frequent words*. Gambar 2 merupakan bentuk arsitektur dari *Continuous Bag of Word (CBOW)*.



Gambar 2. *Continuous Bag of Word (CBOW)*

2) *Skip-Gram*

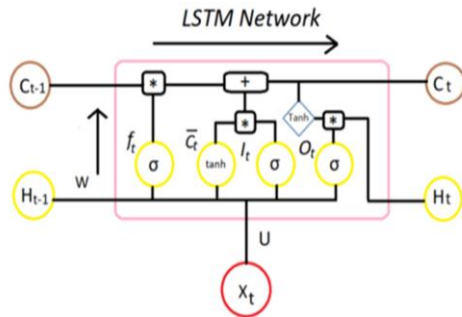
Skip-Gram merupakan model yang menggunakan kata untuk memprediksi target konteks. *Skip-Gram* bekerja dengan baik walaupun dengan data pelatihan yang jumlahnya sedikit dan model ini dapat merepresentasikan kata-kata yang dianggap langka. Gambar 3 merupakan arsitektur dari *Skip-Gram*.



Gambar 3. *Skip-Gram*

D. Pembuatan Model Menggunakan LSTM

Setelah *word embedding* menggunakan Word2Vec proses selanjutnya adalah klasifikasi menggunakan algoritma *Long Short-Term Memory (LSTM)*. Metode *deep learning LSTM* dipilih karena beberapa pertimbangan antara lain LSTM memiliki akurasi yang baik dalam mengolah data berupa teks, LSTM merupakan pengembangan dari metode *deep learning RNN* yang memiliki kelebihan mampu memproses data yang relatif panjang serta LSTM merupakan algoritma yang sangat efisien karena LSTM memiliki fitur komponen ekstra pada sel yang dimiliki oleh *basic recurrent unit*. Komponen LSTM terbagi menjadi beberapa bagian antara lain *Memory Cell*, *Input Gate*, *Forget Gate*, *Output Gate*. Struktur LSTM dapat dilihat pada Gambar 4.



Gambar 4. Struktur LSTM

Keterangan Gambar 4:

- X_t = Input Vector
- H_{t-1} = Previous Cell Output
- C_{t-1} = Previous Cell Memory
- H_t = Current Cell Output
- C_t = Current Cell Memory
- W, U = Weight vector for forget gate (f), candidate (c), i/p gate (l) and o/p gate (O)

Proses komputasi pada LSTM dilakukan dengan beberapa tahapan. Nilai dari suatu *input* hanya dapat disimpan ke dalam *cell state* hanya jika diijinkan oleh *input gate*. Perhitungan dari nilai pada *input gate* dan kandidat dari *cell state* dilakukan dengan menggunakan rumus (2) dan (3).

$$l_t = \sigma (X_t * U_i + H_{t-1} * W_i) \quad (2)$$

$$\bar{C}_t = \tanh(X_t * U_c + H_{t-1} * W_c) \quad (3)$$

Dimana l_t Merupakan keluaran dari *input gate* yang mengontrol seberapa besar informasi dari sel memori yang diusulkan (\bar{C}_t) dan akan diperbarui dengan informasi baru dari *input* saat ini (X_t). \bar{C}_t merupakan *candidate memory cell* pada saat ini, dihitung dengan menggunakan fungsi \tanh dari hasil perkalian *input* (X_t) dengan bobot (U_c) dan keluaran sel sebelumnya (H_{t-1}) dengan bobot (W_c). \tanh digunakan untuk menghasilkan nilai antara -1 dan 1. H_{t-1} Merupakan keluaran dari sel sebelumnya (atau biasa disebut sebagai “*hidden state*”) dari *timestep* sebelumnya. U_f, U_c, U_i, U_o Merupakan matriks bobot yang digunakan untuk mengubah *input* (X_t) menjadi nilai-nilai yang akan memengaruhi masing-masing *forget gate*, *memory cell*, *input gate* dan *output gate*. W_f, W_c, W_i, W_o Merupakan matriks bobot (*weight*) yang digunakan untuk mengubah keluaran (H_{t-1}) menjadi nilai-nilai yang akan memengaruhi masing-masing *forget gate*, *memory cell*, *input gate*, dan *output gate*. Kemudian nilai dari *fore gate* dihitung dengan menggunakan rumus (4).

$$f_t = \sigma (X_t * U_f + H_{t-1} * W_f) \quad (4)$$

Dimana f_t merupakan *output* dari *forget gate*. *Gate* ini mengendalikan sejauh mana informasi dari (C_{t-1}) sel LSTM atau unit GRU sebelumnya yang akan dilupakan atau dipertahankan pada *Current Cell Memory* (C_t). Nilai yang digunakan antara 0 dan 1 untuk menentukan berapa banyak informasi yang akan dilupakan dan (σ) merupakan fungsi

sigmoid yang digunakan untuk menghasilkan nilai antara 0 dan 1 sebagai hasil dari perhitungan rumus yang ada di dalam tanda kurung. Fungsi *sigmoid* umumnya digunakan untuk menghasilkan nilai probabilitas. Selanjutnya *memory cell state* dihitung menggunakan rumus (5).

$$C_t = f_t * C_{t-1} + l_t * \bar{C}_t \quad (5)$$

Dimana C_t merupakan sel memori saat ini yang diperbarui dengan menggunakan hasil dari *forget gate* (f_t), keluaran dari *input gate* (l_t), dan sel memori yang diusulkan (\bar{C}_t). Setelah dihasilkan *memory cell state* yang baru, nilai dari *output gate* dapat dihitung dengan menggunakan rumus (6).

$$O_t = \sigma (X_t * U_o + H_{t-1} * W_o) \quad (6)$$

Dimana O_t merupakan keluaran dari *output gate* yang mengendalikan seberapa besar informasi dari *Current Cell Memory* (C_t) akan digunakan untuk menghasilkan *output* (H_t). Nilai *output final* dihitung menggunakan rumus (7) sebagai berikut :

$$H_t = O_t * \tanh (C_t) \quad (7)$$

Dimana H_t merupakan keluaran saat ini dari sel, yang dihasilkan dengan mengalikan keluaran dari *output gate* (O_t) dengan \tanh dari *Current Cell Memory* (C_t).

Dalam membuat model LSTM struktur dasar dari algoritma untuk membangun dan melatih model LSTM terdapat pada Algoritma 1.

Algoritma 1 Pseudocode LSTM
1. <i>Input</i> : $x = (x_1, \dots, x_{365}), x_i, \in \mathbb{R}^n$
2. <i>Dengan Parameter</i> : $W_f, U_f, W_c, U_c, W_i, U_i, W_o, U_o$
3. <i>Inisialisasi</i> $h_o, c_o = \vec{0}$
4. <i>for</i> $t = 1, \dots, 365$ <i>do</i>
5. Hitung f_t (Eq. 2), \bar{c}_t (Eq. 3), l_t (Eq. 4)
6. Update cell state c_t (Eq. 5)
7. Hitung o_t (Eq. 6), h_t (Eq. 7)
8. <i>end for</i>
9. <i>Output</i> : $h = (h_1, \dots, h_{365}), h_i, \in \mathbb{R}^m$

Algoritma LSTM digunakan dalam berbagai tugas yang melibatkan pemrosesan data sekuensial dan LSTM dirancang untuk mengatasi masalah pengingat informasi jangka panjang dalam data sekuensial.

IV. HASIL EKSPERIMEN DAN ANALISIS

Dalam tahap ini penulis membahas hasil eksperimen dan analisis model yang dibangun. Hasil eksperimen dan analisa terdiri dari pengumpulan data, *preprocessing*, penggunaan model *word embedding* Word2Vec, penggunaan model LSTM, pengujian model dan *output* hasil analisa sentimen.

A. Dataset

Dataset diambil dengan cara *scrapping* menggunakan *Apify Instagram Scraper* mulai dari tanggal 9 Januari 2019 sampai dengan tanggal 30 Januari 2023. Data yang didapat sebanyak 1.506 *caption* dan 4.637 opini. Data *caption* dikelompokkan kedalam lima bagian bidang. Pembagian bidang tersebut merupakan pengelompokan instansi sesuai dengan fungsinya yang antara lain bidang pembangunan, bantuan sosial, layanan masyarakat, perindustrian, dan pendidikan. Hasil dari pengelompokan bidang dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pembagian Bidang

Bidang	Jumlah
Bidang Bantuan Sosial	408
Bidang Pembangunan	293
Bidang Perindustrian	282
Bidang Pendidikan	265
Bidang Layanan Masyarakat	258
Total	1506

Data opini masyarakat yang diperoleh diberi label untuk menentukan sentimen opininya yang antara lain positif, negatif atau netral. *Dataset* yang sudah diberi label akan dimuat oleh sistem untuk dilakukan proses *preprocessing*, *word embedding*, pembuatan model LSTM, pelatihan dan pengujian *output* hasil sentimen opini publiknya. Hasil jumlah pelabelan data dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pelabelan Data

Opini	Jumlah
Netral	1575
Negatif	1588
Positif	1474
Total	4637

B. Data Preprocessing

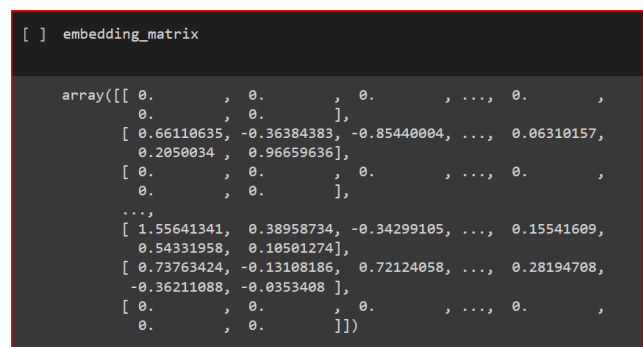
Dataset yang dikumpulkan dalam penelitian ini merupakan data yang tidak berstruktur, sehingga perlu dilakukan *preprocessing* untuk membersihkan data tersebut. Proses tersebut dilakukan agar data yang dihasilkan dapat membuat hasil klasifikasi lebih optimal. Hasil dari tahap *preprocessing* dapat dilihat pada Tabel 3.

C. Word2vec

Dalam Penelitian ini menggunakan *library gensim* dalam membuat model Word2Vec yang telah dilatih sebelumnya pada data Bahasa Indonesia. Model yang digunakan dalam penelitian ini menggunakan metode *Continuous Bag of Word* (CBOW) yang dimuat dari *file* "idwiki_cord2vec_300.model". Metode tersebut dipilih agar waktu training dapat lebih cepat dijalankan dan memiliki akurasi yang baik dalam proses *training model*. Sebelum proses membuat *embedding*, data *preprocessing* diubah menjadi *sequences* dan diberi *padding* dengan panjang maksimal 300. Gambar 5 hasil *embedding matrix* dalam bentuk *array*.

Tabel 3. Hasil Tahap Preprocessing

Proses	Komentar Asli	Hasil Proses
Case Folding	Tolong supaya menjadi perhatian.jalan jaksa agung suprapto 1 sisi jalan TIDAK ADA SELOKAN NYA,mohon supaya di buatkan selokan untuk daerah jaksa agung suprapto,terima kasih 🙏	tolong supaya menjadi perhatian.jalan jaksa agung suprapto 1 sisi jalan tidak ada selokan nya,mohon supaya di buatkan selokan untuk daerah jaksa agung suprapto,terima kasih 🙏
Removal URL	Mampir saja ke sini Gaes https://ayoberbagi.id/wisata-edukasi-batik-bojonegoro-di-desa-ringintunggal	Mampir saja ke sini Gaes
Remove Punctuation	Tanda larangan dan jalur 1 arah tdk berfungsi (banyak yg melanggar)	Tanda larangan dan jalur 1 arah tdk berfungsi banyak yg melanggar
Remove Hashtag	Mungkin bisa dilakukan verifikasi ulang atas data yang ada, apakah penilaiannya dulu benar2 sesuai parameter atau tidaknya. Karena berdasarkan pengamatan di sekitar saya sebenarnya sangat sedikit yang benar2 tergolong miskin. #untuk meminimalkan gejala yang punya rumah bagus, sawah luas, dll bisa mendapatkan bantuan.	Mungkin bisa dilakukan verifikasi ulang atas data yang ada, apakah penilaiannya dulu benar2 sesuai parameter atau tidaknya. Karena berdasarkan pengamatan di sekitar saya sebenarnya sangat sedikit yang benar2 tergolong miskin. untuk meminimalkan gejala yang punya rumah bagus, sawah luas, dll bisa mendapatkan bantuan.
Remove Mention	Kerenn.. tidak hanya Insentif Marbot/Takmir Masjid, Guru Madin/TPQ, Modin Perempuan, dll. Tapi hingga RT/RW 🙏	Kerenn.. tidak hanya Insentif Marbot/Takmir Masjid, Guru Madin/TPQ, Modin Perempuan, dll. Tapi hingga RT/RW 🙏 pemkabbojonegoro @pembkabbojonegoro
Convert Emoticon	Ibu, Mau tanya yang CAR CSR? 🙏🙏	Ibu, Mau tanya yang CAR CSR? Tangan Berdoa Wajah Tertawa
Convert Word	Alhamdulillah.. smga kekurangan guru di SD kami segera terisi. 🙏🙏🙏	Alhamdulillah.. semoga kekurangan guru di SD kami segera terisi. 🙏🙏🙏
Remove Stopword	Program untuk guru Sekolah Minggu dong Min... Kan guru TPQ sudah dapat insentif juga 🙏 biar seimbang dan tidak berat sebelah...	Program untuk guru Sekolah Minggu Min... Kan guru TPQ sudah dapat insentif juga 🙏 biar seimbang tidak berat sebelah...
Stemming	SMPN 3 KEDUNGADEM Keindahan gerak musik dan kostum sangat serasi 🙏 Sehingga gerakan tari nya Begitu kompak dan mempesona 🙏	SMPN 3 KEDUNGADEM indah gerak musik dan kostum sangat serasi 🙏 Sehingga gerak tari begitu kompak dan pesona 🙏



Gambar 5. Embedding Matrix Word2Vec

D. LSTM

Dalam membuat model LSTM struktur dasar dalam membangun dan melatih model LSTM menggunakan pustaka *deep learning* “keras” yang berjalan diatas “TensorFlow”. Dalam penelitian ini ditambahkan 32-unit lapisan (*neuron*) kedalam model lapisan LSTM. Aktivasi yang digunakan dalam sel LSTM adalah “*tanh*” dan pengaturan *regularisasi* (0,01) untuk mencegah *overfitting*. Dalam pembuatan model juga ditambahkan lapisan tambahan *dropout* dan *dense*. Lapisan *dropout* secara acak menonaktifkan sebagian unit (*neuron*) yang terdapat pada lapisan sebelumnya selama pelatihan, sehingga setiap iterasi dalam pelatihan beberapa unit diabaikan dan model harus belajar dengan cara yang lebih tahan terhadap *overfitting*. Lapisan *dense* merupakan *layer* yang digunakan sebagai *output layer*, *dense layer* merupakan sebuah lapisan dimana setiap unit (*neuron*) terhubung dengan unit (*neuron*) pada lapisan berikutnya. *Sequential* digunakan untuk menggabungkan berbagai *layer* secara berurutan. Dalam penelitian ini menggunakan *softmax* dalam klasifikasi *multiclass* dan “*adam*” *compile* untuk menentukan *optimizer*.

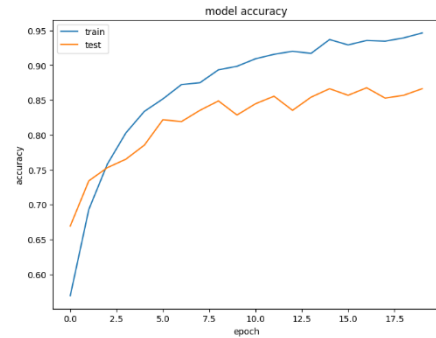
E. Pelatihan Model

Setelah proses klasifikasi dengan menggunakan LSTM, proses selanjutnya adalah pelatihan model. Parameter yang digunakan dalam pelatihan model dapat dilihat pada Tabel 4.

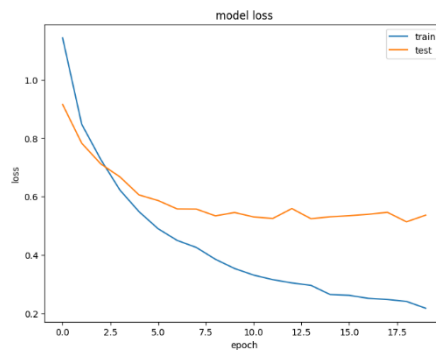
Tabel 4. Parameter

Parameter	Jumlah
<i>K-Fold</i>	5
<i>Batch Size</i>	8
<i>Dropout Rate</i>	0,5
<i>Embedding Dim</i>	300
<i>Max Sequence Length</i>	300
<i>Epoch</i>	20
<i>Learning Rate</i>	0,001

Pelatihan Model dalam penelitian ini menggunakan *K-Fold Cross-Validation*, dimana merupakan teknik yang digunakan untuk mengevaluasi kinerja sebuah model dengan cara yang lebih *reliable*. Teknik ini membantu mengatasi masalah *overfitting* dan memungkinkan agar mendapat perkiraan yang lebih baik dalam model berkinerja pada data yang belum pernah dilihat sebelumnya [21]. Dimulai dari pembagian data yang menjadi lima bagian yang setara, lalu salah satu dari pembagian akan diiterasi digunakan sebagai data *testing* sementara pembagian data lainnya digunakan sebagai data *training*. Pada setiap iterasi terdapat proses melatih model yang digunakan sebagai data pelatihan dan menjadi *Fold 1* dan *Fold* yang tersisa menjadi data uji yang tidak digunakan untuk pelatihan. Hasil akurasi klasifikasi menggunakan LSTM dalam pelatihan model pada *epoch* ke 20, validasi *Accuracy Curves* sudah diatas 80% dan pada grafik validasi *Loss Curves* nilai *loss* dari data *train* semakin kecil. Berikut Gambar 6 merupakan grafik model *accuracy* dan Gambar 7 model *loss* yang dihasilkan selama pelatihan model yang dibuat.



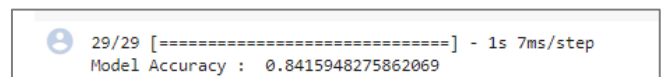
Gambar 6. Model Accuracy



Gambar 7. Model Loss

F. Pengujian Model

Setelah model selesai dilatih dengan parameter yang telah ditentukan, langkah selanjutnya adalah pengujian model. Pengujian model bertujuan untuk melihat sejauh mana model tersebut berhasil dalam melakukan tugas yang ditargetkan dalam klasifikasi sentimen. Dalam mengetahui performa sebuah model diperlukan sebuah evaluasi, salah satu metode evaluasi adalah *confusion matrix*. *Confusion Matrix* dapat diterapkan pada klasifikasi biner serta masalah klasifikasi *multiclass* [22]. Dalam *confusion matrix* terdapat beberapa matrik yang antara lain *accuracy*, *recall*, *precision* dan *F1-score*. *Accuracy* digunakan untuk mengukur sejauh mana model dapat mengklasifikasi data dengan benar secara keseluruhan. Hasil akurasi yang didapat dari model yang dilatih menghasilkan akurasi sebesar 84,16% seperti terlihat pada Gambar 8.



Gambar 8. Accuracy Model

Recall digunakan untuk mengukur sejauh mana model dapat mendeteksi hasil positif yang ada. *Precision* digunakan untuk mengukur sejauh mana model dapat menghindari membuat prediksi positif yang salah. *F1-Score* merupakan matrik gabungan antara *precision* dan *recall* untuk memberikan pemahaman keseluruhan tentang kinerja model. Hasil dari evaluasi data bidang dapat dilihat pada Tabel 5 dan hasil dari evaluasi data opini dapat dilihat pada Tabel 6 dibawah ini:

Tabel 5. Hasil Evaluasi Data Bidang

Bidang	Precision	Recall	F1-Score
Pendidikan	66.7%	73.7%	70%
Layanan Masyarakat	77.8%	87.5%	82.4%
Perindustrian	83.3%	88.2%	85.7%
Pembangunan	89.7%	83.9%	86.7%
Bantuan Sosial	95.9%	94.2%	95%

Tabel 6. Hasil Evaluasi Data Opini

Opini	Precision	Recall	F1-Score
Positif	86.7%	74.6%	80.2%
Negatif	82.6%	92.5%	87.2%
Netral	83.7%	85.4%	84.6%

Confusion matrix juga digunakan untuk memvisualisasikan kinerja prediksi model pada *supervised learning*. Setiap data dari masing-masing kelas dalam tabel *confusion matrix* menunjukkan jumlah prediksi yang dibuat untuk mengklasifikasi kelas yang benar atau salah. Pada penelitian ini menggunakan klasifikasi *confusion matrix* tiga kelas. Hasil pengujian dapat dilihat pada Gambar 9 dimana hasil sentimen positif sebanyak 250, netral sebanyak 246 dan negatif sebanyak 288.



Gambar 9. Confusion Matrix

Setelah model selesai dibangun dan dilakukan evaluasi selanjutnya dilakukan pengujian model. Dalam proses pengujian model data uji dilakukan proses *preprocessing* kembali untuk dibersihkan datanya. Data hasil *preprocessing* diberi token dan dilakukan prediksi terhadap bidang dan sentimennya. Hasil prediksi bidang dan sentimen opini publik dapat dilihat pada Gambar 10 dan prosentase prediksi opini pada setiap bidang untuk mengetahui dalam opini masyarakat lebih banyak yang positif, netral atau negatif dapat dilihat pada Tabel 7.

text	Prediksi Bidang	Bidang %	Prediksi Sentiment	Sentiment %
0 jalan banyak yang rusak	Pembangunan	99.54%	Negatif	97.48%

Gambar 10. Hasil Uji Coba Klasifikasi Sentimen Opini Publik

Tabel 7. Hasil Uji Coba Sentimen Pada Masing-Masing Bidang

Bidang	Netral	Positif	Negatif
Pendidikan	25.89%	43.99%	30.12%
Pembangunan	29.39%	39.06%	31.54%
Bantuan Sosial	19.46%	42.62%	37.92%
Layanan Masyarakat	32.03%	43.26%	24.71%
Perindustrian	31.87%	37.85%	30.28%

V. KESIMPULAN

Penelitian klasifikasi sentimen opini publik pada akun resmi Instagram Pemerintah Kabupaten Bojonegoro menggunakan *word embedding* Word2Vec dan algoritma LSTM memberikan hasil akurasi 84,16%. Model *word embedding* Word2Vec yang dimuat dari "idwiki_cord2vec_300.model" dapat menyingkat waktu *training*. Metode *deep learning* LSTM dipilih karena memiliki akurasi yang baik dalam mengolah data berupa teks. Model *word embedding* Word2Vec dan algoritma LSTM dapat menjadi salah satu pilihan untuk analisa sentimen opini publik. Hasil analisa tersebut dapat memberikan kontribusi positif dan dapat menjadi bahan pertimbangan Pemerintah Kabupaten Bojonegoro dalam upaya meningkatkan layanan masyarakat, program dan kebijakan yang dibuat.

REFERENSI

- [1] G. D'Aniello, M. Gaeta, and I. La Rocca, "KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis," *Artif Intell Rev*, vol. 55, no. 7, pp. 5543–5574, 2022.
- [2] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, pp. 1–33, 2017.
- [3] Y. Chen, J. Yuan, Q. You, and J. Luo, "Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 117–125.
- [4] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *eProceedings of Engineering*, vol. 2, no. 1, 2015.
- [5] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment analysis using Word2vec and long short-term memory (LSTM) for Indonesian hotel reviews," *Procedia Comput Sci*, vol. 179, pp. 728–735, 2021.
- [6] W. Suciska, "Optimalisasi Penerapan E-Government melalui Media Sosial dalam Mewujudkan Good Governance," in *Prosiding Seminar Nasional Komunikasi" Akselerasi Pembangunan Masyarakat Lokal Melalui Komunikasi dan Teknologi Informasi*", 2016, pp. 374–389.
- [7] N. Y. A. Faradhillah, R. P. Kusumawardani, and I. Hafidz, "Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin," *SESINDO 2016*, vol. 2016, 2016.
- [8] "Pengguna Media Sosial di Indonesia Capai 191 Juta pada 2022." [Online]. Available: <https://dataindonesia.id/digital/detail/pengguna-media-sosial-di-indonesia-capai-191-juta-pada-2022>

- [9] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 11, pp. 4704–4713, 2018.
- [10] A. F. Hidayatullah and A. S. N. Azhari, "Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter," in *Seminar Nasional Informatika (SEMNASIF)*, 2015.
- [11] M. Qasem, R. Thulasiram, and P. Thulasiram, "Twitter sentiment classification using machine learning techniques for stock markets," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 834–840.
- [12] I. Kurniasari, K. Kusriani, and H. Al Fatta, "Analysis of Public Opinion Sentiment on Instagram regarding Covid-19 with SVM," *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem dan Komputer*, vol. 1, no. 1, pp. 67–74, 2021.
- [13] M. J. Naufal, D. Richasdy, and M. A. Bijaksana, "Sentiment Analysis of Student Satisfaction on Telkom University Language Center (LaC) Services on Instagram Using the RNN Method," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 181–190, 2022.
- [14] A. M. Rafi, S. Rana, R. Kaur, Q. M. J. Wu, and P. M. Zadeh, "Understanding global reaction to the recent outbreaks of covid-19: Insights from instagram data analysis," in *2020 IEEE international conference on systems, man, and cybernetics (SMC)*, 2020, pp. 3413–3420.
- [15] G. Borah, D. Nimje, G. JananiSri, K. P. Bharath, and M. R. Kumar, "Sentiment Analysis of Text Classification Using RNN Algorithm," in *Proceedings of International Conference on Communication and Computational Technologies: ICCCT 2021*, 2021, pp. 561–571.
- [16] M. A. Rahman, H. Budianto, and E. I. Setiawan "Aspect Based Sentimen Analisis Opini Publik Pada Instagram dengan Convolutional Neural Network" *INSYST*, vol. 1, no. 2, pp. 50–57, Dec. 2019, doi: 10.52985/insyst.v1i2.83.
- [17] F. A. Nugraha, N. H. Harani, and R. Habibi, *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Kreatif, 2020.
- [18] Syaiful Imron, E. I. Setiawan, Joan Santoso, and Mauridhi Hery Purnomo, "Aspect Based Sentiment Analysis Marketplace Product Reviews Using BERT, LSTM, and CNN," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 586–591, Jun. 2023, doi: 10.29207/resti.v7i3.4751.
- [19] H. Liu, "Sentiment analysis of citations using word2vec," *arXiv preprint arXiv:1704.00177*, 2017.
- [20] A. Nurdin, B. Anggo, S. Aji, A. Bustamin, and Z. Abidin, "Perbandingan Kinerja Word Embedding Word2vec, Glove, Dan Fasttext Pada Klasifikasi Teks," *Jurnal TEKNOKOMPAK*, vol. 14, no. 2, p. 74, 2020.
- [21] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 5, pp. 577–584, 2018.
- [22] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," in *Data democracy*, Elsevier, 2020, pp. 83–106.