

**Pen-O-Matis: Penulis Konten Otomatis Menggunakan Metode  
*Recurrent Neural Network* (RNN) untuk Membantu  
Pemasaran Digital UMKM pada Masa Pandemi**



**Dosen Pembimbing:**  
**Muhammad Fachrie, S.T., M.Cs.**

**Disusun oleh:**

|                            |                   |
|----------------------------|-------------------|
| <b>Agung Wirakurniawan</b> | <b>5180411340</b> |
| <b>Maulana Kavaldo</b>     | <b>5180411365</b> |
| <b>Alfian Hidayatulloh</b> | <b>5180411382</b> |

**PROGRAM STUDI S-1 INFORMATIKA  
FAKULTAS SAINS & TEKNOLOGI  
UNIVERSITAS TEKNOLOGI YOGYAKARTA**

## ABSTRAK

Pesatnya perkembangan teknologi kecerdasan buatan di seluruh dunia juga selaras dengan semakin diperlukannya digitalisasi dalam kehidupan sehari-hari, terutama setelah pandemi Covid-19 menyerang ke seluruh sektor perekonomian. Namun hal tersebut ternyata belum mampu diikuti oleh para pelaku UMKM di Indonesia. Pada akhirnya, banyak UMKM yang terpaksa menutup tokonya karena tidak mampu bersaing di era yang serba digital. Oleh sebab itu, penulis mengajukan sebuah penelitian tentang *text generation* (penghasil teks), yang merupakan salah satu penerapan teknologi kecerdasan buatan di mana komputer dapat menuliskan sebuah teks berupa kalimat secara otomatis berdasarkan input kata kunci tertentu. *Text Generation* ini memiliki kemampuan menghasilkan teks secara alami (*natural language*) sehingga menyerupai teks yang ditulis oleh manusia secara langsung. Teknologi ini sangat cocok diterapkan ke dalam sistem untuk membantu para pelaku UMKM dalam membuat konten-konten di media sosial secara otomatis. Untuk dapat menghasilkan sistem penghasil teks secara otomatis, membutuhkan model dengan metode *Recurrent Neural Network* (RNN) yang memiliki sistematika perhitungan bobot secara berulang. RNN merupakan salah satu metode *Deep Learning* di mana nilai akurasi yang dihasilkan dapat lebih baik dibandingkan dengan jaringan syaraf tiruan sederhana. Jenis arsitektur RNN yang digunakan adalah LSTM (*Long Short-Term Memory*) karena dapat menutupi kekurangan pada RNN yang tidak dapat menyimpan memori untuk dipakai kembali sebagai input pembelajaran. *Dataset* yang digunakan adalah data hasil *web scraping* ke beberapa akun Instagram *food blogger* di Indonesia dengan jumlah sebanyak 1.540 data. Penelitian ini menguji performa sistem menggunakan metode *Human Evaluation* yaitu menilai tata bahasa dan relevansi kalimat yang dihasilkan secara otomatis terhadap kalimat dalam bahasa alami manusia.

**Kata kunci:** *Text Generation, Recurrent Neural Network, LSTM, Natural Language Generation Model, Konten Media Sosial*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Penyebaran *Coronavirus disease* (Covid-19) telah menimbulkan berbagai dampak yang signifikan terhadap perekonomian di Indonesia. Pandemi yang telah berlangsung lebih dari satu tahun ini menyerang hampir seluruh sektor usaha, tak terkecuali UMKM. Walhasil banyak pengusaha mengalami penurunan pendapatan, yang kemudian mengakibatkan beberapa UMKM terpaksa gulung tikar. Berdasarkan data Kementerian Koperasi dan Usaha Kecil Menengah (Kemenkop UKM) bersama Asosiasi UMKM Indonesia (Akumindo), UMKM pada tahun 2017 berjumlah 60,4 juta dan terus meningkat menjadi 62,6 juta pada 2018 dan 64,7 juta pada 2019. Namun, jumlah UMKM pada 2020 menurun drastis menjadi 34 juta unit (Kemenkop UKM, 2020).

Era transformasi digital ditandai dengan transaksi *e-commerce* pada kuartal II-2020 yang naik hampir 40% menjadi 383 juta, dibanding dengan kuartal I-2020 sebesar 275 juta (Bank Indonesia, 2021). Hal ini menunjukkan bahwa masyarakat yang tadinya tidak berbelanja *online*, sudah mulai menysar *e-commerce*. Meski digitalisasi terbuka lebar, nyatanya kendala masih sering dialami oleh pelaku UMKM dalam memasuki dunia digital. Kurangnya pengetahuan usaha secara online, bingung untuk memasarkan produk melalui platform digital, hingga kesiapan tenaga ahli yang dimiliki, membuat banyak pelaku UMKM terpaksa tetap berjualan secara *offline*.

Beragam upaya telah dilakukan oleh pemerintah untuk terus mempertahankan dan meningkatkan daya beli masyarakat ke UMKM. Salah satunya yakni berkolaborasi dengan *marketplace* tanah air untuk memberikan edukasi mengenai pentingnya UMKM *go digital*. Di tengah penyebaran wabah Covid-19 yang belum mereda, memanfaatkan pemasaran UMKM melalui media sosial dan *marketplace*, dirasakan cukup efektif untuk mengurangi adanya kontak langsung dengan pelanggan. Meskipun sudah mengenal dan menggunakan platform digital untuk berjualan *online*, beberapa pelaku UMKM mengaku belum mampu memaksimalkan penjualan karena terbatasnya ide untuk membuat konten media sosial. Untuk itu, diperlukan sebuah sistem yang dapat membantu mereka dalam membuat konten-konten (teks) digital secara gratis dan mudah.

Melalui pemanfaatan kecerdasan buatan, permasalahan yang dialami oleh para pelaku UMKM tersebut dapat teratasi berkat teknologi yang bernama *machine learning*. Teknologi ini memungkinkan mesin (komputer) untuk berpikir selayaknya manusia dengan algoritma-algoritma yang terus menerus dikembangkan seiring berjalannya waktu. Selain itu, dengan memanfaatkan *natural language processing* (NLP), nantinya mampu menciptakan sistem yang bisa memahami bahasa manusia sehari-hari. Dengan dasar-dasar di atas, penulis mengajukan penggunaan *deep learning* dan NLP (*text generation*) dengan metode *recurrent neural network* untuk dapat menghasilkan konten-konten media sosial secara otomatis. Penulis juga berharap, penelitian ini dapat membantu para pelaku UMKM di Indonesia agar bisa bertahan menghadapi masa-masa sulit saat pandemi.

## **1.2 Batasan Penelitian**

Batasan masalah yang ditentukan untuk menghindari perluasan pembahasan dalam penelitian ini adalah sebagai berikut:

1. Konten-konten media sosial yang dihasilkan secara otomatis oleh model terbatas pada format teks, bukan gambar maupun suara.
2. Penelitian ini terbatas pada tahap perancangan model dengan metode *recurrent neural network*, sehingga belum sampai pada tahap implementasi ke dalam sistem.

## **1.3 Tujuan Penelitian**

Adapun tujuan yang ingin dicapai oleh penulis pada penelitian ini meliputi:

1. Mendapatkan model yang mampu menghasilkan konten-konten media sosial secara otomatis, terutama konten berbasis teks.
2. Mengetahui performa model penulis konten media sosial secara otomatis yang menggunakan metode *recurrent neural network* (RNN).

## **1.4 Manfaat Penelitian**

Pada penelitian ini, terdapat beberapa manfaat yang akan diperoleh berdasarkan tujuan penelitian antara lain:

1. Dari sisi akademis, untuk mengetahui performa model RNN dalam *text generation* sehingga bisa menjadi pembanding pada penelitian-penelitian selanjutnya.
2. Dari sisi sosial, untuk membantu para pelaku UMKM di Indonesia dalam memasarkan produk secara digital sehingga bisa memberikan kesejahteraan ekonomi bagi keluarga dan masyarakat sekitar.

## BAB II

### TINJAUAN PUSTAKA

Beberapa variasi model *text generation* menggunakan LSTM telah banyak dikembangkan oleh peneliti-peneliti di seluruh dunia. Santhanam, S. (2020) membangun model yang mampu menghasilkan teks berdasarkan masukan serangkaian kata berbentuk vektor konteks. Vektor konteks mirip dengan vektor paragraf yang mengambil makna semantik (konteks) dari suatu kalimat. Dataset yang digunakan dalam penelitiannya berupa skrip dialog dari film *Lord of The Rings* sebanyak 100.000 kalimat. Arsitektur model yang digunakan adalah LSTM (*Long Short-Term Memory*) dengan konfigurasi berbasis *bidirectional* LSTM, 256 *hidden units*, fungsi aktivasi ReLU untuk *input layer* dan fungsi aktivasi Softmax untuk *output layer*. Dalam pengujiannya, digunakan metode evaluasi *Cosine similarity measure* yang didasarkan pada hipotesis bahwa kata benda dalam kalimat memuat sebagian besar konteks dibandingkan dengan jenis kata lain dalam kalimat.

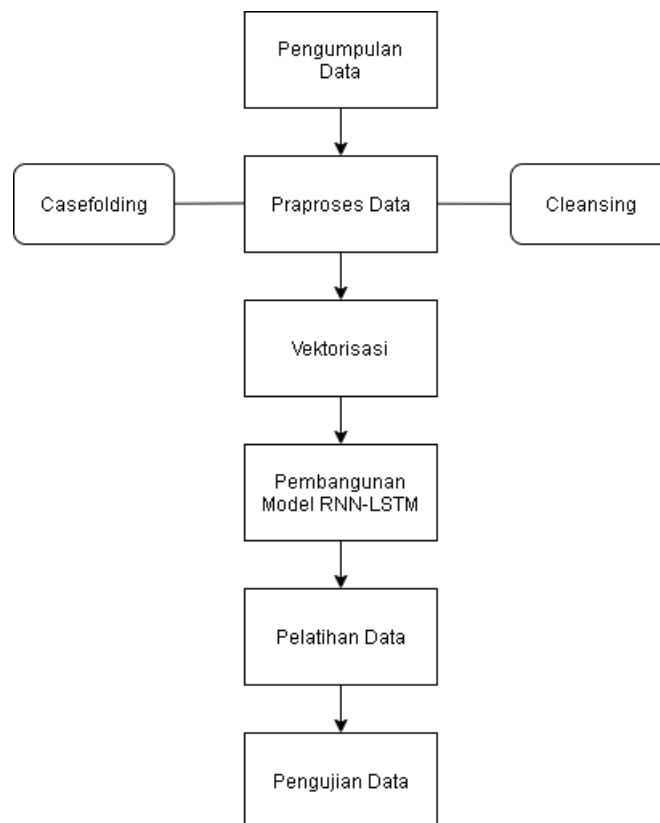
Penelitian lain yang dilakukan oleh Pawade, D., dkk. (2018) menggunakan metode RNN-LSTM dengan konfigurasi yang berbeda-beda untuk menghasilkan teks berupa cerita atau kalimat naratif. Dataset yang digunakan sebagai input pelatihan yaitu cerita dari film *Alice in Wonderland* dan *Charlie and the Chocolate Factory* yang mana berisi sekitar 5000 kosa kata. Evaluasi performa pada model ini menggunakan metode *Human Evaluation* yang mana mengukur tata bahasa, keterkaitan kalimat yang dihasilkan dengan cerita, nilai dari pembaca, dan keunikan dalam cerita yang dihasilkan.

Berbeda dengan dua penelitian sebelumnya, Cao, J., dkk. (2018) mengimplementasikan metode LSTM untuk menghasilkan model yang mampu mentranslasi teks berbahasa Mandarin ke bahasa Inggris. Dataset yang digunakan merupakan hasil *web scraping* ke media sosial Weibo sebanyak 60.000 kalimat dengan kata kunci Kobe Bryant. Selain itu, juga dilakukan *data preprocessing* menggunakan *library* khusus bernama “jieba” agar dapat menyesuaikan dengan huruf-huruf yang ada dalam bahasa Mandarin. Untuk menguji model hasil pelatihan, digunakan metode *Perplexity* yang mengukur kualitas bahasa alami dan *LSI similarity* yang mengukur relevansi kata yang dihasilkan oleh model dengan topik bahasan. Dalam pengujiannya, peneliti juga membandingkan model terhadap model lain yaitu HMM dan Char-RNN.

## BAB III

### METODOLOGI PENELITIAN

Tahapan-tahapan yang dilaksanakan pada penelitian ini ditunjukkan pada Gambar 3.1. Tahap pertama adalah pengumpulan data, tahap kedua praproses data, tahap ketiga proses vektorisasi, tahap keempat proses pemodelan menggunakan metode RNN-LSTM, tahap kelima proses pelatihan data, dan tahap keenam adalah proses pengujian data.



**Gambar 3.1** Diagram alur penelitian

#### 3.1 Perangkat Uji Coba

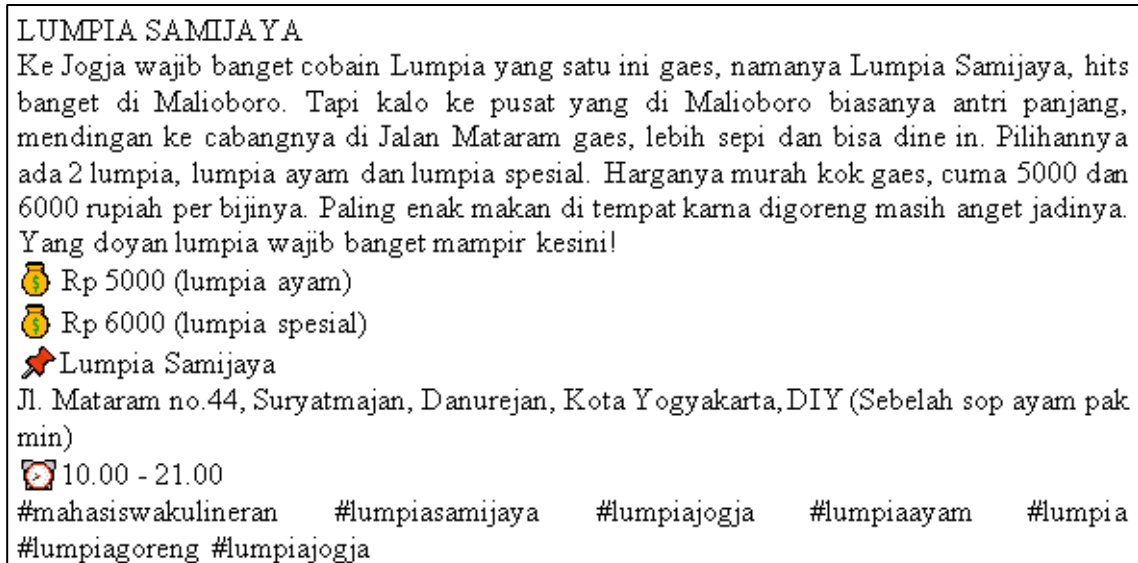
Dalam proses penelitian, penulis menggunakan perangkat lunak Google Colaboratory, yaitu layanan berbasis *cloud* yang memungkinkan pengguna untuk menulis dan mengeksekusi bahasa pemrograman Python melalui *browser*. Adapun spesifikasi perangkat keras dan lunak yang digunakan dalam layanan tersebut meliputi:

1. *Library* dan API:
  - Scrapy versi 2.5.0
  - API versi 2.0.3, *powered by* WebScraping.AI
  - Pandas versi 1.1.5

- RE versi 2.2.1
  - String dan OS (built-in) versi 3.7.11
  - IPython-Autotime versi 0.3.1
  - Tensorflow dan Keras versi 2.5.0
  - Numpy versi 1.19.5
  - Matplotlib versi 3.2.2
2. Sistem operasi: Linux
  3. Prosesor: Intel Xeon Processor 2.3 GHz (1 Core, 2 Thread)
  4. RAM: 12 GB
  5. SSD: 68 GB
  6. GPU: Nvidia Tesla K80
  7. Bahasa pemrograman: Python versi 3.7.11

### 3.2 Pengumpulan Data

Studi kasus pada penelitian ini bertujuan untuk menghasilkan konten media sosial berbasis teks. Untuk memperoleh data yang berupa korpus berbahasa Indonesia, penulis menggunakan cara *web scrapping* pada beberapa akun Instagram milik *food blogger* Indonesia. Adapun akun-akun Instagram yang penulis gunakan sebagai sumber data adalah @mahasiswakulineran, @anakjajan, @jogjakakilima, dan @jogjafooddestinations. Akun tersebut dipilih karena berisi konten-konten yang mewakili UMKM bertema kuliner di Indonesia. Salah satu contoh takarir (*caption*) dari akun Instagram @mahasiswakulineran dapat dilihat pada Gambar 3.1 di bawah ini.



**Gambar 3.2** Contoh takarir akun Instagram @mahasiswakulineran

### 3.3 Praproses Data

Data yang diperoleh dari proses *web scraping* tentunya tidak hanya memuat karakter berupa huruf dan angka. Oleh karena itu, penulis juga melakukan praproses data untuk normalisasi dan menghilangkan *noise*. Praproses data yang dilakukan antara lain *case folding* (menyeragamkan huruf kapital), serta menghapus karakter non-ASCII, *emoticon*, *hashtag*, URL, *whitespace*, dan angka. Dataset kemudian akan dilakukan tokenisasi sehingga menghasilkan jumlah kosa kata sebanyak 12.915 kata. Kosa kata ini selanjutnya akan masuk ke proses vektorisasi untuk dikonversi ke bentuk data vektor.

### 3.4 Vektorisasi

Pada tahap ini, kata yang sudah dibersihkan dan siap digunakan diubah terlebih dahulu menjadi vektor (angka) dengan tujuan agar tiap kata memiliki nilai perbedaan antara satu dengan yang lain. Vektorisasi pada penelitian ini menggunakan metode *word embedding*, yaitu dengan pemberian nilai indeks pada tiap kata sehingga efektif saat dilakukan perhitungan data dan menjadi mudah diidentifikasi oleh sistem.

### 3.5 Pembangunan Model

Proses pembangunan model pada penelitian ini memerlukan data korpus (teks) berbahasa Indonesia yang penulis peroleh dari hasil *web scraping* yang memuat sebanyak 1.540 takarir. Pembuatan model dilakukan dengan melatih dataset melalui skenario *hyperparameters tuning*. *Hyperparameters tuning* adalah proses pemilihan dan pengoptimalan berbagai hyperparameter di dalam jaringan syaraf tiruan, termasuk LSTM. Pemilihan kombinasi beberapa hyperparameter dapat menghasilkan model dengan performa yang berbeda-beda. Proses pembuatan model ini dilakukan dengan menyesuaikan berbagai parameter seperti jumlah *layer*, *recurrent units*, *input size*, *batch size*, *epoch*, maupun *dropout* (Reimers, N. dan Gurevych, I., 2017).

Salah satu jenis *neural network* adalah *recurrent neural network* (RNN). Jaringan ini memiliki sedikit perbedaan daripada jaringan *neural network* biasa. Pada *neural network* biasa, setiap input akan masuk melalui *hidden layer* dan mengeluarkan *output*, sedangkan pada RNN terdapat perulangan. Jadi setiap input yang masuk dan menghasilkan *output*, akan masuk kembali menjadi input untuk diproses di dalam *hidden layer* (Graves, A., 2012). Begitu seterusnya sehingga memperoleh *output* yang sesuai dengan target. Model dibangun dengan metode RNN secara *bidirectional* atau perhitungan bobot pada RNN secara bolak-balik. Jenis arsitektur RNN yang digunakan adalah LSTM yang



memiliki kelebihan daripada sekedar RNN biasa, yaitu adanya arsitektur mengingat dan melupakan *output* yang akan diproses kembali menjadi input. Selain itu, kemampuan LSTM yang lain adalah dapat mempertahankan akurasi yang terjadi ketika melakukan *backpropagation* sehingga mencegah *loss rate* meningkat (Fauzi, R., 2016).

### 3.6 Pelatihan Data

Implementasi pada proses pelatihan dilakukan dengan menggunakan data latih yang sudah ada. Hasil proses pelatihan adalah sebuah model yang mampu menghasilkan teks berupa kalimat berdasarkan sebuah input kata kunci tertentu. Pada *encoder layer*, sel LSTM akan memproses input sehingga menghasilkan *output* yang akan diteruskan ke *hidden layer*. Pada *hidden layer*, data yang dilatih akan dicatat dalam bentuk logistik untuk memperbaharui bobot dan nilai yang dibutuhkan agar mencapai target *output*. *Decoder layer* akan menerima hasil dari *hidden layer* untuk kemudian difokuskan menggunakan *attention* sehingga mendapatkan bobot nilai yang mendekati nilai target *output*. Lalu bentuk vektor akan diubah kembali menjadi teks untuk kemudian ditampilkan sebagai *output* teks. Demi mendapatkan hasil yang optimal, diperlukan beberapa nilai parameter pada proses pelatihan data seperti pada Tabel 3.1 berikut:

**Tabel 3.1** Nilai parameter pada pelatihan data

| Parameter                               | Nilai                            |
|---|----------------------------------|
| Jumla data latih                        | 1.540 takarir, 12.915 kata       |
| Jumlah <i>max features</i>              | 10.000 kata                      |
| Ukuran <i>sequence input</i>            | 4 kata                           |
| Ukuran <i>sequence output</i>           | 1 kata                           |
| Ukuran <i>buffer</i> pada TF.shuffle    | 1.000                            |
| Ukuran <i>buffer</i> pada TF.prefetch   | dinamis                          |
| Ukuran <i>batch</i>                     | 32                               |
| Ukuran <i>embedding</i>                 | 128                              |
| Jumlah LSTM <i>units</i>                | 128                              |
| Jumlah <i>hidden layer</i>              | 3                                |
| Jumlah <i>epoch</i>                     | 20                               |
| Algoritma optimasi penyesuaian bobot    | Adam                             |
| Algoritma untuk fungsi <i>loss rate</i> | Sparse Categorical Cross Entropy |
| Fungsi aktivasi untuk prediksi kata     | Softmax                          |
| Ukuran kata kunci untuk data uji        | 2 frasa, 8 kata                  |

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Hasil Proses Web Scraping

Setelah melakukan *web scraping* pada beberapa akun Instagram *food blogger* di Indonesia menggunakan *library* Scrapy dan layanan API gratis, diperoleh dataset sebanyak 1.540 takarir. Adapun salah satu contoh data (takarir) hasil *web scraping* di akun Instagram @mahasiswakulineran dapat dilihat pada Gambar 4.1 di bawah ini.

| captions  |
|---|
| MAKANAN BALI DI JOGJA\n\nMakan malam di hari minggu ini yang enak enak dulu gaes, tapi ini NON HALAL ya gaes, yang gak berkenan skip aja hehehe\n\nNah ini makannya di Bu Kadek Khas Bali yang ada di Selokan Mataram bagian Kalasan Sleman\n\nDisini pesen 3 menu sekaligus, nasi campur spesial, sate B2, dan sate lilit B2. Harganya terjangkau kok gaes semuanya.\n\nNasi campur spesialnya mantap, dapet nasi trancam, B2 Rica, B2 Kecap, sate B2, lawar, serta kuah sup. Rasanya bikin nagih, nggak cukup 1 porsi hahaha\n\nSate Lilit B2 nya gede gede, seporsi dapat 5 tusuk. Bisa buat lauk ataupun digadoin. Rasanya manis gurih dan bikin nagih juga, muantap\n\nSate B2 nya request pedes , bumbunya meresap ke dalam dagingnya, pedesnya masih enak nggak terlalu pedes\n\nOverall semuanya recommended untuk dicoba\n\n📌 Menu di #pricelistkuliner_bukadekkhasbali📌 Bu Kadek Khas Bali\nJl. Selokan Mataram, Kadirojo I, Purwomartani, Kalasan, Sleman, DIY\n📅 Selasa - Minggu🕒 09.00 - 21.00\n\n#mahasiswakuliner... |

Gambar 4.1 Contoh data hasil *web scraping*

#### 4.2 Hasil Proses Data Preprocessing

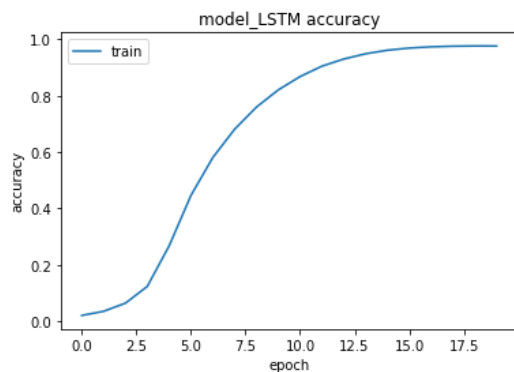
Dari hasil *web scraping*, diketahui bahwa masih ada banyak *noise* atau karakter-karakter yang kurang penting dalam sebuah kalimat. Oleh karena itu, penulis melakukan *data preprocessing* dengan menormalisasi huruf kapital (*case folding*) dan membersihkan kalimat dari simbol, *emoticon*, atau karakter yang tidak diperlukan (*cleansing*) sehingga menghasilkan sebuah kalimat seperti pada Gambar 4.2 berikut.

| captions  |
|---|
| makanan bali di jogja makan malam di hari minggu ini yang enak enak dulu gaes tapi ini non halal ya gaes yang gak berkenan skip aja hehehe nah ini makannya di bu kadek khas bali yang ada di selokan mataram bagian kalasan sleman disini pesen menu sekaligus nasi campur spesial sate dan sate lilit harganya terjangkau kok gaes semuanya nasi campur spesialnya mantap dapet nasi trancam rica kecap sate lawar serta kuah sup rasanya bikin nagih nggak cukup porsi hahaha sate lilit nya gede gede seporsi dapat tusuk bisa buat lauk ataupun digadoin rasanya manis gurih dan bikin nagih juga muantap sate nya request pedes bumbunya meresap ke dalam dagingnya pedesnya masih enak nggak terlalu pedes overall semuanya recommended untuk dicoba menu di bu kadek khas bali bu kadek khas bali jl selokan mataram kadirojo purwomartani kalasan sleman diy selasa minggu |

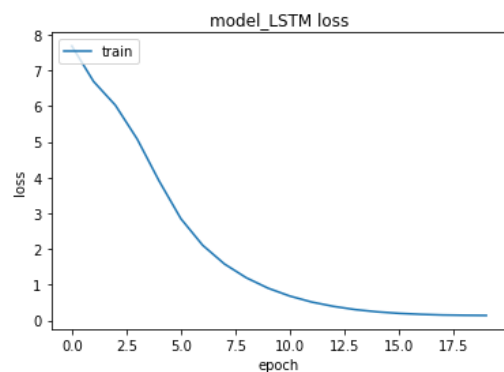
Gambar 4.2 Contoh data hasil *preprocessing*

### 4.3 Hasil Proses Pelatihan Data

Pada proses pelatihan data, digunakan dua pengukuran untuk melihat kualitas dari hasil proses pelatihan, yaitu nilai akurasi dan nilai *loss*. Nilai akurasi pada proses pelatihan cenderung meningkat seiring meningkatnya jumlah *epoch* yang dilakukan, di mana pada *epoch* ke-20, nilai akurasi mencapai 97,63%. Untuk nilai *loss*, dapat dilihat bahwa semakin banyak epoch yang dilakukan, nilai *loss* juga semakin menurun. Perbandingan grafik dari proses pelatihan data seperti terlihat pada Gambar 4.3 dan Gambar 4.4.



**Gambar 4.3** Nilai akurasi pelatihan



**Gambar 4.4** Nilai *loss* pelatihan

### 4.4 Hasil Proses Pengujian Data

Untuk menguji performa dari model yang telah diperoleh pada proses pelatihan, penulis menggunakan data sampel kata kunci yang sering diketikkan pada kolom mesin pencari Google. Adapun data sampel kata kunci dan hasil dari model penulis konten media sosial (berbasis teks) secara otomatis seperti terlihat di bawah ini.

#### 1) Data uji ke-1

- Kata kunci  
“nasi goreng pedas enak”
- Hasil *text generation*  
“nasi goreng pedas enak karna cuma yang nyaman kalian beli di coffee km hotel biru jl kaliurang km aesthetic pondok sepatu menanti seharga flaky khusus ice of dessert suasana ini bisa dimakan makan disini rp siomay rp bakso rp tusuk bakmi puas rp bakmi goreng bakmi dengan babi dipadu babi yang apa dengan harga nya cuma terjangkau nya ada btw kali ini buka di spot outlet dengan saus dibikin yang mereka punya indomie jadi aku sate kambing yang harganya ati harga nya udah gudeg gaes tanpa photo banget namanya gofood nih gaes dengan harga nasi campur rp termasuk nasi jumbo es teh nasi goreng”

## 2) Data uji ke-2

- Kata kunci  
“warung cafe kopi favorit”
- Hasil text generation  
“warung cafe kopi favorit di jogja follow spot kuliner bikin kalian homemade dan tunggu banget terjangkau juga tau buka dari warung store story tapi ini good kamu udah buat kalian makan di sei dessert sei sapi dalam terakhir kuah dulu lagi lagi mereka lagi ada promo hanya nih di shopeetokopedia space yang kulineran ajak yogyakarta belanja dan maranggi grab no pengen maguwoharjo solo harga wib durian jogjakakilimaofficial jogja yang martabak ylokita kembangan floo setoran temen kalian only spot coffeeshop di durian cabang komplek ini kamu buka puasa menu tanpa boba selamat khusus tgl maret goodluck khusus berlaku id tag temen kalian penggemar haii iphoneplast”

Metode evaluasi yang penulis gunakan pada penelitian ini adalah *Human Evaluation*. Parameter yang digunakan sebagai penilaian meliputi tata bahasa (X1), keterkaitan kalimat yang dihasilkan dengan kata kunci (X2), nilai dari pembaca (X3), dan keunikan dalam teks yang dihasilkan (X4). Untuk rentang nilai atau skor yang diterapkan antara 0 – 100. Untuk rentang kesimpulan terhadap total nilai meliputi sangat baik (A), baik (B), cukup baik (C), buruk (D), dan sangat buruk (E). Hasil proses pengujian data yang dilakukan oleh penulis seperti terlihat pada Tabel 4.1 sebagai berikut.

**Tabel 4.1** Hasil proses pengujian data

| Data uji            | Parameter Penilaian |    |    |    | Total      | Akurasi |
|---------------------|---------------------|----|----|----|------------|---------|
|                     | X1                  | X2 | X3 | X4 |            |         |
| Data 1              | 43                  | 51 | 45 | 54 | 193        | 48,25%  |
| Data 2              | 67                  | 70 | 69 | 65 | 271        | 67,75%  |
| Rata-rata Penilaian |                     |    |    |    | 232        | 58%     |
| Kesimpulan Penulis  |                     |    |    |    | Cukup Baik |         |

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Setelah melakukan pengujian dan evaluasi terhadap penghasil teks otomatis, ada beberapa poin yang dapat disimpulkan antara lain:

1. Metode *Recurrent Neural Network* dengan arsitektur LSTM dapat digunakan untuk membangun model penulis konten media sosial secara otomatis dengan cukup baik, sehingga mendekati kemiripan terhadap teks dalam bahasa alami manusia.
2. Dengan menggunakan metode *Human Evaluation*, diperoleh informasi bahwa performa model penulis konten media sosial otomatis ini cukup baik.

#### **5.2 Saran**

Adapun saran yang dapat diberikan oleh penulis setelah melakukan penelitian ini terkait perbaikan dan pengembangan untuk penelitian selanjutnya adalah sebagai berikut:

1. Sebelum melakukan *web scraping*, direkomendasikan membeli paket ke penyedia layanan API *web scraping* agar bisa memperoleh jumlah API call yang cukup.
2. Saat melakukan *data preprocessing*, penulis menyarankan untuk dilakukan proses *spelling correction* dan translasi bahasa asing pada hasil *web scraping* agar bisa menghasilkan kualitas *dataset* yang lebih baik.
3. Saat melakukan pelatihan data, disarankan memaksimalkan proses *hyperparameter tuning* agar bisa mendapatkan model yang akurat dan mampu menuliskan kalimat yang menyerupai dengan teks dalam bahasa alami manusia.
4. Untuk penelitian selanjutnya, penulis berharap model yang telah dirancang sebaiknya diimplementasikan ke dalam sebuah situs web atau aplikasi supaya para pelaku UMKM di Indonesia bisa mencoba langsung dan merasakan manfaatnya.

## DAFTAR PUSTAKA

- Bank Indonesia. (2021), *Laporan Tahunan Bank Indonesia Tahun 2020*, [PDF], (<https://www.bi.go.id/id/publikasi/laporan/Documents/Laporan-Akuntabilitas-Bank-Indonesia-2020.pdf>), Diakses 28 Juli 2021.
- Cao, J., & Wang, C., (2018), *Social Media Text Generation Based on Neural Network Model*, Proceedings of the 2nd International Conference on Computer Science and Artificial Intelligence - CSAI '18.
- Fauzi, R., (2016), *Implementasi Jaringan Syaraf Tiruan dengan Metode Backpropagation terhadap Bibit Tanaman Karet*. Jurnal Education and Development STKIP Tapanuli Selatan, 1-11.
- Graves, A., (2012), *A Comparison of Network Architectures. In Supervised Sequence Labelling with Recurrent Neural Networks*, Heidelberg: Springer, 47-56.
- Kementerian Koperasi dan Usaha Kecil Menengah, (2020), *Perkembangan Data Usaha Mikro, Kecil, Menengah (UMKM) dan Usaha Besar (UB) Tahun 2018 – 2019*, [PDF], ([https://www.kemenkopukm.go.id/uploads/laporan/1617162002\\_SANDINGAN\\_DATA\\_UMKM\\_2018-2019.pdf](https://www.kemenkopukm.go.id/uploads/laporan/1617162002_SANDINGAN_DATA_UMKM_2018-2019.pdf)), Diakses 28 Juli 2021.
- Pawade, D., Sakhapara, A., Jain, M., Jain, N., & Gada, K., (2018), *Story Scrambler - Automatic Text Generation using Word Level RNN-LSTM*, International Journal of Information Technology and Computer Science (IJITCS), 10(6), 44-53.
- Reimers, N. dan Gurevych, I., (2017), *Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks*, arXiv, DOI: arXiv.1707.06799.
- Santhanam, S., (2020), *Context-based Text Generation using LSTM Networks*, arXiv, DOI: arXiv.2005.00048.