

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Subjek dan Objek Penelitian**

Subjek merupakan data yang akan diamati dan diolah selama penelitian berlangsung. Pada penelitian ini menggunakan subjek berupa sekumpulan teks cerita novel yang disatukan ke dalam suatu file txt. Objek penelitian ini yaitu pembuatan teks cerita otomatis menggunakan algoritma *Recurrent Neural Network* dengan hasil yang diharapkan berupa model yang dapat memprediksi kata selanjutnya berdasarkan masukan kata.

#### **3.2. Alat dan Bahan Penelitian**

Terdapat dua macam *instrument* yang digunakan sebagai penunjang berjalannya penelitian ini, diantaranya:

##### 1. Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan untuk pemodelan penelitian adalah laptop Asus VivoBook A409JP dengan spesifikasi sebagai berikut :

- a. *Processor* : Intel Core i5 10<sup>th</sup> Gen
- b. *Graphic Card* : NVIDIA GeForce MX330
- c. *Memory* : 4 GB RAM
- d. *Storage* : 1TB & 512 GB SSD
- e. *Operating System* : Windows 10

##### 2. Perangkat Lunak (*Software*)

Perangkat lunak yang untuk mengembangkan pemodelan, yaitu :

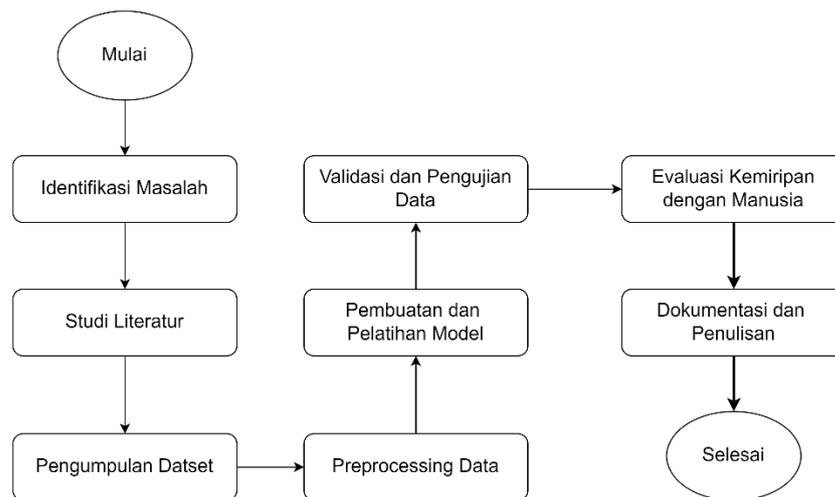
- a. Notepad++
- c. Google Collaboratory
- d. Python
- e. Google Form
- f. Google Spreadsheet

### 3. Bahan

Adapun bahan yang digunakan dalam penelitian ini, yaitu *dataset* teks cerita novel. Kumpulan cerita novel dalam format pdf disatukan seluruhnya. Setelah itu, diubah ke dalam file teks berformat txt agar memudahkan proses *preprocessing* data.

### 3.3. Diagram Alir Penelitian Penelitian

Pelaksanaan penelitian ini mengacu pada diagram alir yang telah dibuat. Hal tersebut berupaya agar penelitian dilaksanakan secara berurutan berdasarkan tahapan atau proses yang ada. Gambar 3.1 menampilkan diagram alir penelitian.



Gambar 3.1 Diagram Alir Penelitian

#### 3.3.1 Identifikasi Masalah

Pelaksanaan penelitian dimulai dari identifikasi masalah yang ada dan ingin diselesaikan. Topik penelitian ini mengambil topik mengenai pembuatan teks cerita otomatis. Topik ini dibahas karena belum adanya kajian yang membahas *text generate* berbahasa Indonesia. Identifikasi masalah dilakukan dengan melakukan pencarian informasi di internet, sehingga didapatkan bahwa menulis cerita membutuhkan waktu untuk melimpahkan

perasaan dan imajinasi ke dalam penulisan. Tidak jarang para penulis mengalami hambatan dalam memikirkan kelanjutan ide dari cerita yang dibangun. Oleh karena itu, dibutuhkannya pembuatan teks cerita otomatis yang dapat merekomendasikan kalimat dari kata yang dimasukkan oleh *user* yang dapat dijadikan sebagai bahan pertimbangan dalam menulis kelanjutan cerita.

### 3.3.2 Studi Literatur

Tahap selanjutnya adalah menentukan studi literatur dengan mencari sumber-sumber yang berkaitan dengan permasalahan yang memiliki kemiripan dengan penelitian yang akan dilaksanakan untuk menambah wawasan terkait. Studi literatur dilakukan dengan mencari dan membaca dokumen dalam bentuk buku, skripsi, jurnal dan *website*. Dokumen-dokumen yang digunakan sebagai referensi tersebut memiliki keterkaitan dengan novel, *text generate*, *Recurrent Neural Network*, *Long-Short-Term Memory*, dan *Gated Recurrent Unit*. Menetapkan studi literatur membantu untuk mendapatkan sebuah informasi terkait kemiripan objek, atau teknik penelitian, sehingga dapat membantu dalam meningkatkan kualitas program dengan pengembangan baru.

### 3.3.3 Pengumpulan Dataset

Tahap berikutnya adalah pengumpulan data berupa teks cerita novel karangan Tere Liye dalam Series Bumi dengan 15 judul novel berbeda dan teks novel karangan J.K Rowling dalam Series Harry Potter dengan 7 judul buku berbeda yang telah diterjemahkan ke dalam Bahasa Indonesia. Data yang digunakan tersebut merupakan data sekunder yang diambil dari salah satu platform penyedia *e-book*, yaitu Z-Library (<https://z-lib.is/>). Dataset tersebut digabungkan ke dalam satu file dengan panjang teks sebanyak 1.943.055 kata dengan format txt. Penggalan dataset tersebut dapat dilihat pada Tabel 3.1.

Tabel 3.1 Dataset Novel

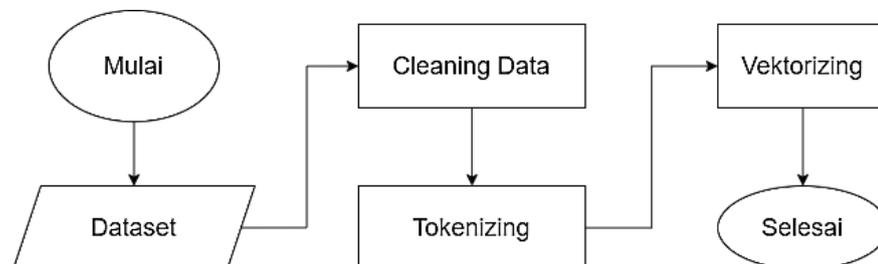
Judul	Isi
<i>Harry Potter and The Goblet of Fire</i>	PENDUDUK desa Little Hangleton masih menyebutnya "Rumah Riddle", meskipun sudah bertahun-tahun lamanya keluarga Riddle tak tinggal di sana lagi. Rumah itu terletak di atas bukit, menghadap ke desa, beberapa di antara jendela-jendelanya ditutup papan,
<i>Harry Potter and The Deathly Hallows</i>	Dua orang itu muncul secara tiba-tiba, terpisah beberapa meter di sebuah jalan sempit yang diterangi oleh cahaya bulan. Sesaat mereka berdiri diam, tongkat masing-masing saling terarah ke dada yang lain. Setelah mengenali satu sama lain,
<i>Harry Potter and The Chamber of Secrets</i>	BUKAN untuk pertama kalinya pertengkaran meledak di meja makan rumah Privet Drive nomor empat. Sebelumnya Mr Vernon Dursley telah terbangun pagi-pagi buta oleh bunyi keras dari kamar keponakannya, Harry
<i>Harry Potter and The Sorcerer's Stone</i>	Mr. dan Mrs. Dursley yang tinggal di Privet Drive nomor empat bangga menyalakan din bahwa mereka orang-orang yang normal, untunglah. Mereka tak bisa diharapkan terlibat dengan sesuatu yang ajaib atau misterius,
<i>Harry Potter and The Prisoner of Azkaban</i>	HARRY POTTER adalah anak yang sangat istimewa dalam banyak hal. Misalnya saja, dia paling benci liburan musim panas dibanding waktu-waktu lainnya. Contoh lain lagi, dia ingin sekali mengerjakan PR-nya,
<i>Harry Potter and The Order Of Phoenix</i>	Hari terpanas sejauh ini pada musim panas telah mulai berakhir dan keheningan yang membuat mengantuk melanda rumah-rumah besar berbentuk bujursangkar di Privet Drive. Mobil-mobil yang biasanya mengkilat diliputi debu di jalan-jalan
<i>Harry Potter and The Half Blood Prince</i>	Saat itu menjelang tengah malam dan perdana Menteri sedang duduk sendirian di kantornya, membaca laporan panjang yang lewat begitu saja melalui otaknya tanpa meninggalkan makna sedikit pun.
Hujan	RUANGAN 4 x 4 m2 itu selintas terlihat didesain terlalu sederhana untuk sebuah ruangan paling mutakhir di kota ini. Padahal ruangan itu berteknologi tinggi dan berperalatan medis paling maju. Teknologi terapinya tidak pernah dibayangkan manusia sebelumnya.
Bumi	NAMAKU Raib. Aku murid baru di sekolah. Usiaku lima belas tahun. Aku anak tunggal, perempuan. Untuk remaja se-umuranku, tidak ada

Judul	Isi
	yang spesial tentangku. Aku berambut hitam, panjang, dan lurus.
Bulan	GERIMIS membungkus halaman sekolah. Langit mendung. Gumpalan awan hitam seakan bosan beranjak di atas sana. Satu-dua tetes air mengenai jendela kelas lalu terbawa angin. Udara terasa lembap dan dingin. Ini sebenarnya sudah di ujung musim hujan.
Matahari	PUKUL satu siang. Hujan turun deras di luar. Suara petir terdengar susul- menyusul, angin kencang berkesiur. Udara terasa lembap dan dingin.
Bintang	ALI, sejujurnya Bapak tidak punya ide sama sekali apa yang sebenarnya kamu lakukan setiap kali menghadapi kertas ulangan: Pak Gun menatap dari balik kaca mata tebalnya. "Entah kenapa nilai-nilai ulanganmu selalu saja buruk.
Ceroz dan Batazoar	PAGI hari pukul 07.30. Suara lantang panggilan menaiki pesawat terdengar di langit-langit bandara. Aku segera memasukkan buku ke tas ranselku, memastikan tidak ada yang tertinggal, menyiapkan boarding pass pesawat, lalu beranjak berdiri.
Komet	Musim hujan telah datang. Gerimis membungkus pagi, terus turun tak berkesudahan sejak tadi malam. Jendela mobil berembun, udara terasa dingin.
Komet Minor	Permukaan laut semakin mendidih. Buih memenuhi sekitar. Hanya soal detik sekarang, portal menuju klan Komet Minor akan terbuka. Max alias Si Tanpa Mahkota alias penipu itu, berdiri menunggu di atas Pulau Hari Minggu
Selena	Namaku Selena. Aku lahir di Distrik Sabit Enam, dua ratus kilometer utara Kota Tishri, Klan Bulan. Itu bukan kawasan yang maju dan canggih. Itu kawasan kumuh, padat dan tertinggal.
Nebula	Libur panjang berjalan seru dan menyenangkan. Setiba di rumah Paman Raf, aku menyaksikan bangunan lima lantai itu terlihat bersolek, dengan lampu hias warna-warni, itu persiapan acara pernikahan Am empat minggu lagi.
Lumpu	Basemen besar rumah Ali lengang sejenak setelah sambungan komunikasi antar klan dunia paralel itu terputus. Tidak ada lagi layar yang menunjukkan Miss Selena, atau Lumpu. Juga tidak terlihat lagi ruangan bebatuan yang basah, dengan hewan pengerat berkeliaran.

Judul	Isi
Si Putih	Cerita ini bermula di Klan Polaris. Sebuah Klan yang unik di konstelasi Ursa. Hari itu, tanggal 1, minggu 1, bulan 1, tahun 1, dalam sistem kalender Klan Polaris. Kalian tidak keliru membacanya, itu memang hari pertama penanggalan mereka, tapi itu bukan berarti klan itu baru ada, usia klan itu jauh lebih tua dibanding konstelasi lain.
Bibi Gill	“Kalau aku jadi kamu, aku tidak akan melakukannya, anak muda!” Seseorang telah muncul di depan Raja yang menunggangi Naga. Seseorang itu mengambang di udara.
Sagaras	“Kurang ini, Lae.” Sopir angkot menggeleng. “Eh, kurang apanya? Segitu kan cukup buat berdua.” Seli bertanya balik. “Tarif angkot naik, Neng. Masa' kamu tidak tahu sih? Kamu ke mana saja? Menurut Surat Keputusan Wali Kota nomor sekian-sekian.” Sopir angkot menjelaskan, menyebut berapa kurangnya.
Matahari Minor	Halo, namaku Seli. Kalian tentu sudah mengenalku, bukan? Aku remaja usia tujuh belas, kelas sebelas. Aku sama seperti Sekolah, belajar, jalan-jalan ke mal, suka makan yang enak-enak, suka baca buku, nonton film, suka menonton drama Korea, juga suka boy band. Ibuku dokter, ayahku pegawai kantoran. Aku anak tunggal, jadi seluruh kasih sayang dan perhatian orangtua tumpah kepadaku.

### 3.3.4 Preprocessing Data

Selanjutnya pada tahapan *preprocessing* data. Terdapat tiga tahapan pada *preprocessing* data seperti pada Gambar 3.2, yaitu *cleaning*, *tokenizing* dan *vectorizing*.



Gambar 3.2 *Preprocessing Data*

### 3.3.4.1 Cleaning Data

*Dataset* yang telah disatukan seluruhnya ke dalam file berformat txt, merupakan *dataset* kotor. Hal tersebut berarti masih terdapat banyak karakter tidak penting yang membuat cerita tidak seperti naskah aslinya yang membentuk satu kesatuan kalimat yang lugas. Oleh karena akan dilakukan *cleaning text* pada *dataset* menggunakan *library regex* untuk menghilangkan unsur atau karakter yang tidak penting. Gambar 3.3 menampilkan script yang digunakan untuk *cleaning dataset* serta fungsinya yang pada Tabel 3.2.

```
c_text = re.sub('\- ', '', cobain)
c_text = re.sub('\r\n', '', c_text)
c_text = re.sub('\r\n+', ' ', c_text)
c_text = re.sub('\xad +', '', c_text)
c_text = re.sub('\-|\-|\n+|\'', ' ', c_text)
c_text = re.sub('[ ]+', ' ', c_text)
c_text = re.sub('[^A-Za-z0-9.,!?"" ]', '', c_text)
c_text
```

Gambar 3.3 *Cleaning Dataset*

Tabel 3.2 Fungsi *Syntax Regex*

<b><i>Syntax Regex</i></b>	<b><i>Kegunaan</i></b>
<code>\-</code>	Mecocokkan karakter '-'
<code>\r\n</code>	Mencocokkan <i>newline</i> pada data
<code>\xad</code>	Mencocokkan karakter ' <i>Soft Hyphen</i> '
<code>\n+</code>	Mencocokkan satu atau lebih <i>newline</i>
<code>[ ]</code>	Mecocokkan setiap karakter yang tidak termasuk dalam spasi tunggal
<code>[^A-Za-z0-9.,!?"" ]</code>	Mecocokkan setiap karakter yang tidak termasuk dalam rentang huruf besar (A-Z), huruf kecil (a-z), digit (0-9), tanda baca (.,!?"" ) dan spasi

Hasil *cleaning* data lebih jelasnya seperti yang ditunjukkan pada Tabel 3.3.

Tabel 3.3 Hasil *Cleaning Dataset*

<i>Input</i>	<i>Ouput</i>
selalu tenang pun kehilangan kata-kata. Batozar berdiri. "Eh, Master B mau kemana?" Tanya Ali. "Toilet. Di mana toilet Ruangan ini, heh?" "Lurus ke sana, Master B." Ali menunjuk pojok ruangan, "Tapi, bagaimana dengan Ily." "Apanya yang bagaimana? sudah Aku menyelesaikan Sisanya, tugasku. itu urusan kalian. Sekarang urusanku ad sakit perut. Aku butuh toilet." Splash. Batozar telah melesat menuju pojok basemen.	selalu tenang pun kehilangan kata kata. Batozar berdiri. "Eh, Master B mau kemana?" Tanya Ali. "Toilet. Di mana toilet ruangan ini, heh?" "Lurus ke sana, Master B." Ali menunjuk pojok ruangan, "Tapi, bagaimana dengan Ily." "Apanya yang bagaimana? Aku sudah menyelesaikan tugasku. Sisanya, itu urusan kalian. Sekarang urusanku adalah sakit perut. Aku butuh toilet." Splash. Batozar telah melesat menuju pojok basemen.

### 3.3.4.2 *Tokenizing*

Dilanjutkan dengan *tokenizing*, yaitu pemisahan data teks setiap karakter penyusun dari kalimat, sehingga dapat diolah pada proses vektorisasi[39]. Merupakan tahap untuk memotong kalimat menjadi potongan-potongan kata yang sesuai dengan kebutuhan analisis. Pemotongan kalimat dilakukan dengan melihat spasi yang ada pada tiap kalimat sehingga terpisah masing-masing katanya untuk dapat digunakan dalam melakukan analisis sentiment. Contohnya seperti berikut :

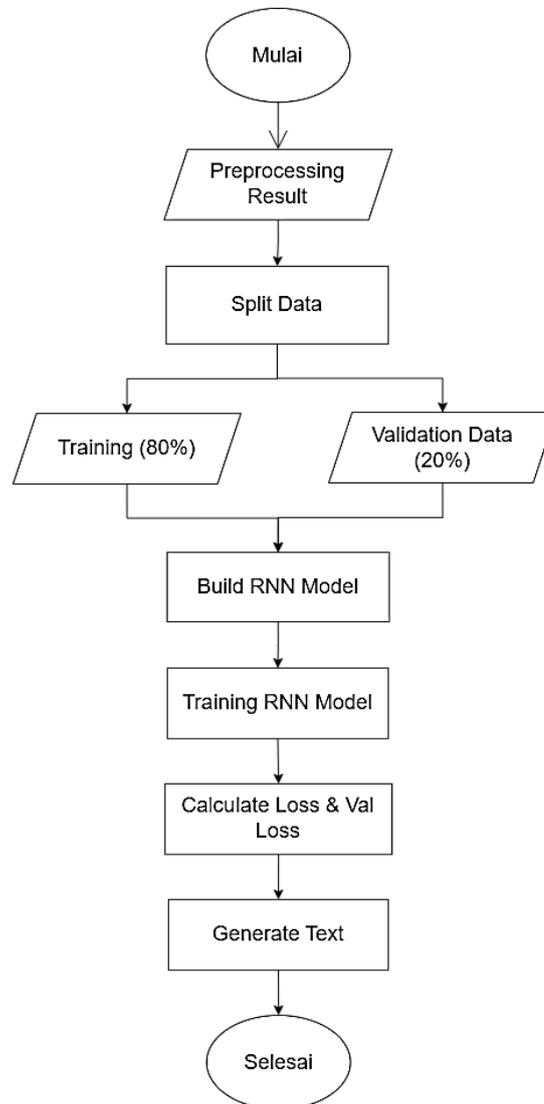
```
['pagi', 'ini', 'cuaca', 'cerah', 'dan'],  
['ini', 'cuaca', 'cerah', 'dan', 'sejuk']
```

### 3.3.4.3 *Vektorizing*

Terakhir data akan masuk pada tahap *vektoring*, yaitu pengubahan huruf menjadi vektor angka yang memudahkan computer untuk membaca data, sebelum nantinya akan dilakukan proses pelatihan atau *training* menggunakan algoritma *Recurrent Neural Network*. Vektorisasi yang dilakukan menggunakan *layer embedding*. *Layer embedding* adalah layer pada model *neural network* yang bertugas untuk mengubah input diskrit seperti kata atau kategori menjadi representasi vektor kontinu. vektor *embedding* dapat menangkap pola-pola yang mendasari hubungan antara kata-kata, sehingga model dapat menghasilkan teks yang lebih bervariasi dan realistis bahkan untuk kata-kata yang tidak muncul dalam data pelatihan.

### 3.3.5 Pembuatan dan Pelatihan Model

Tahapan pelatihan dan pembuatan model digambarkan seperti pada Gambar 3.4.



Gambar 3.4 Pembuatan Model dan *Training* Model

Setelah dataset bersih, akan dilanjutkan dengan pembuatan model untuk melakukan pelatihan dataset dan mengetahui nilai *loss*. Perbandingan *training data* dan *validation data* yang digunakan adalah 80:20[11]. *Training data* digunakan untuk melatih model. *Validation data* digunakan selama pelatihan untuk menguji seberapa baik model belajar dari data latihan tanpa melibatkan data latihan itu

sendiri. Algoritma yang digunakan dalam penelitian ini adalah *Recurrent Neural Network* dengan arsitektur model *Long Short-Term Memory*, dan *Gated Recurrent Unit*. Masing-masing arsitektur dilakukan *training* sebanyak 50 *epochs* dan *batch size* sebesar 50. Terdapat dua jenis model yang dibangun, yaitu 1 *layer* dan 2 *layers*, sehingga total ada 34 model, seperti pada Tabel 3.4 dan Tabel 3.5.

Tabel 3.4 Model 1 *Layer*

<b>1 Layer</b>	
<b>Embedding</b>	<b>Filter</b>
16	16
	32
	64
	128
	256
32	32
	64
	128
	256
64	64
	128
	256

Tabel 3.5 Model 2 *Layers*

<b>2 Layers</b>		
<b>Embedding</b>	<b>Filter</b>	<b>Dropout</b>
32	32, 64	0.2 –
	32, 128	0.2 –
	32, 256	0.2 –

### 3.3.6 Validasi dan Pengujian Data

Parameter yang digunakan dalam perhitungan performa model adalah *embedding*, *loss*, jumlah *layer*, dan jenis arsitektur. Setelah data sudah dilatih kemudian dilakukan proses pengujian data berdasarkan hasil *generate text* algoritma *Recurrent Neural Network*. Dari pelatihan model arsitektur LSTM dan GRU yang telah dilakukan. *Output generate text* terdiri atas 6 kata dari lanjutan kata masukan. Dilakukan perhitungan kata yang relevan dengan kata masukan pada masing-masing *generate* kalimat. Setelah itu, dihitung pula rata-rata relevan dan standar deviasi dari keseluruhan satu model. Hasil tersebut terdapat pada Tabel 3.6 dan Tabel 3.7.

Tabel 3.6 Hasil Perhitungan Model 1 Layer

1 Layer				
Model	GRU		LSTM	
	Rata-Rata	Standar Dev	Rata-Rata	Standar Dev
<b>E16 U16</b>	1,6	2,2706	<b>3,9</b>	<b>1,7288</b>
E16U32	2,9	2,6437	3,8	2,6998
E16U64	2,8	2,3476	3,6	1,6465
E16U128	3,2	2,5734	3,3	2,2632
E16U256	2,1	2,6013	2,2	1,8738
E32U32	2,8	1,7512	1,9	1,5951
E32U64	1,9	2,2828	2,0	2,5386
E32U128	3,0	2,0000	3,1	2,2828
E32 U256	2,5	2,3214	3,1	2,4698
E64 U64	1,9	2,0790	2,2	2,6998
E64U128	3,2	2,4855	3,7	1,7029
E64 U256	3,2	2,1499	3,7	1,9465

Tabel 3.7 Hasil Perhitungan Model 2 Layers

Model	GRU		LSTM	
	Rata-Rata	Standar Dev	Rata-Rata	Standar Dev
U32-64(drop)	1,5	2,1731	1,6	2,0656
U32-64(nodrop)	1,8	2,3476	1,4	1,6465
U32-128(drop)	1,3	1,6364	3,1	2,4698
U32-128(nodrop)	2,7	2,1108	2,8	1,8135
U32-256(drop)	1,4	1,5055	2,0	2,2111
U32-256(nodrop)	1,0	1,4142	1,4	2,0111

Dapat dikatakan bahwa model LSTM *embedding* 16 dan *filter* 16 merupakan model terbaik dengan 3,9 rata-rata kata relevan dan standar deviasi sebesar 1,7288. Semakin tinggi rata-rata dan semakin rendah standar deviasi, maka semakin baik presisi model yang dibangun. Tabel 3.8 yang merupakan hasil generate model tersebut.

Tabel 3.8 Hasil Evaluasi *Output* Model LSTM

LSTM 1LE16U16		
Kata Masukan	Hasil <i>Generate</i> Kalimat	Kata yang Relevan
Pagi ini anak	anak perempuan semua lawan kali lima	4
Pagi ini anak	anak laki laki laki laki laki	3
Dia selalu	menyukai anak anak yang gagal menangkap	6
Dia selalu	berseru, menepis, menaiki tangga, dan mesin	6
bukan kamu	telah pergi. tahu ini." mereka bertanya	2
bukan kamu	masih menjual bola panas ke belakang	6
Coba lihat	belum baik baik suka dengan fantastis,	2
Coba lihat	meski mengambang dengan panik. dia meraih	2
Terdengar suara	yang seperti kembali ke belakang nenek	5
Terdengar suara	dari lantai, lalu berusaha memasang kapsul	3
<b>Rata-Rata</b>		<b>3,9</b>
<b>Standar Deviasi</b>		<b>1,728840331</b>

### 3.3.7 Evaluasi Hasil Oleh Manusia

Terdapat tiga macam evaluasi hasil oleh manusia, yaitu evaluasi kuesioner, perhitungan teorema slovin dan perhitungan distribusi normal. Lebih lengkapnya akan dibahas di bawah ini.

#### 3.3.7.1 Evaluasi Kuesioner

Penelitian ini menggunakan responden dalam pengisian kuesioner *google form* sebanyak 40 orang guru Sekolah Dasar. Kuesioner berisikan penilaian kepuasan terhadap 10 kalimat buatan manusia dan 10 hasil *generate text* model LSTM 1 layer dengan *embedding 16* dan *unit 16*. Model tersebut merupakan model terbaik berdasarkan evaluasi kata yang paling relevan. Rentang penilaian yang digunakan 1 hingga 5, yaitu [14]; 1 = Sangat Tidak Relevan, 2 = Tidak Relevan, 3 = Cukup Relevan, 4 = Relevan, dan 5 = Sangat Relevan. Hasil pengumpulan data kuesioner dapat dilihat pada Tabel 3.9 dan Tabel 3.10.

Tabel 3.9 Hasil Kalimat Buatan Manusia

Kata Masukan	Hasil <i>Generate</i> Kalimat	1	2	3	4	5
Pagi ini anak	dia berkeliling rumah sebagai bentuk antisipasi	6	6	8	12	8
Pagi ini anak	anak katanya ada yang melihat Unta	6	7	9	5	13
Dia selalu	tertarik mendengar komentarku tentang berita yang	4	2	11	11	12
Dia selalu	memberiku es krim. Apa yang terjadi	3	2	12	13	10
bukan kamu	yang harus memutuskan apa yang baik!	4	4	10	8	14
bukan kamu	saja yang punya buku ajaib matematika	2	4	11	8	15
Coba lihat	Seli, dia santai kembali menyendok makananku	5	2	10	14	9
Coba lihat	siapa saja yang duluan berhasil dievakuasi?	0	1	10	10	19
Terdengar suara	berbisik atau tertawa cekikikan dari kejauhan	2	2	9	7	20

<b>Kata Masukan</b>	<b>Hasil <i>Generate Kalimat</i></b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
	Terdengar suara bergemuruh. Ekornya bergerak di dinding batu.	4	3	14	7	12
<b>Jumlah</b>		36	66	312	380	660
<b>Total</b>		1454				
<b>n</b>		400				
<b>Rata-Rata</b>		3,64				
<b>Standar Deviasi</b>		1,27				

Tabel 3.10 Hasil *Generate Text LSTM Embedding 16 Unit 16*

<b>Kata Masukan</b>	<b>Hasil <i>Generate Kalimat</i></b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Pagi ini anak	anak perempuan semua lawan kali lima	9	4	21	2	4
Pagi ini anak	anak laki laki laki laki laki	15	6	9	5	5
Dia selalu	menyukai anak anak yang gagal menangkap	5	10	14	5	6
Dia selalu	berseru, menepis, menaiki tangga, dan mesin	4	5	16	9	6
bukan kamu	telah pergi. tahu ini." mereka bertanya	6	7	13	8	6
bukan kamu	masih menjual bola panas ke belakang	7	9	10	7	7
Coba lihat	belum baik baik suka dengan fantastis,	6	8	13	8	5
Coba lihat	meski mengambang dengan panik. dia meraih	5	5	14	8	8
Terdengar suara	yang seperti kembali ke belakang nenek	3	6	8	13	10
Terdengar suara	dari lantai, lalu berusaha memasang kapsul	7	4	11	9	9
<b>Jumlah</b>		67	128	387	296	330
<b>Total</b>		1208				
<b>n</b>		400				
<b>Rata-Rata</b>		3,02				
<b>Standar Deviasi</b>		1,30				

### 3.3.7.2 Teorema Slovin

Perhitungan evaluasi menggunakan sampel (S) sebanyak 40 guru dan populasi guru Sekolah Dasar di Indonesia (N) sebanyak 1.605.509 guru[40]. Dengan menggunakan teori Slovin, akan didapatkan batas toleransi kesalahan. Sehingga dari perhitungan tersebut didapat margin eror penelitian ini sebesar 15.8%.

$$\varepsilon = \sqrt{\frac{N - S}{NS}} = \sqrt{\frac{1605509 - 40}{1605509 \cdot 40}} = \sqrt{\frac{1605469}{64220360}}$$

$$\varepsilon = 0.1581 = \mathbf{15.8\%}$$

### 3.3.7.3 Distribusi Normal 68-95-99.7 Rule

Selanjutnya dilakukan perhitungan distribusi normal dari hasil evaluasi kuesioner yang telah dilakukan. Perhitungan tersebut untuk mengelompokkan hasil presentase sebaran data diantaranya dari 68%, 95% dan 99.7% pengamatan. Berikut untuk masing-masing perhitungan distribusi normal pada *generate* manusia dan model:

#### 1. Hasil Kalimat Buatan Manusia

##### a. Presentase sebaran data dari 68% pengamatan

$$68\% \rightarrow (\mu - 1\sigma) \leq X \leq (\mu + 1\sigma)$$

$$(3.64 - 1.27) \leq X \leq (3.64 + 1.27)$$

$$2.37 \leq X \leq 4.91$$

##### b. Presentase sebaran data dari 95% pengamatan

$$95\% \rightarrow (\mu - 2\sigma) \leq X \leq (\mu + 2\sigma)$$

$$(3.64 - 2(1.27)) \leq X \leq (3.64 + 2(1.27))$$

$$1.1 \leq X \leq 6.18$$

c. Presentase sebaran data dari 99.7% pengamatan

$$99.7\% \rightarrow (\mu - 3\sigma) \leq X \leq (\mu + 3\sigma)$$

$$(3.64 - 3(1.27)) \leq X \leq (3.64 + 3(1.27))$$

$$-0.17 \leq X \leq 7.45$$

## 2. Hasil *Generate Text Model*

a. Presentase sebaran data dari 68% pengamatan

$$68\% \rightarrow (\mu - 1\sigma) \leq X \leq (\mu + 1\sigma)$$

$$(3.02 - 1.30) \leq X \leq (3.02 + 1.30)$$

$$1.72 \leq X \leq 4.32$$

b. Presentase sebaran data dari 95% pengamatan

$$95\% \rightarrow (\mu - 2\sigma) \leq X \leq (\mu + 2\sigma)$$

$$(3.02 - 2(1.30)) \leq X \leq (3.02 + 2(1.30))$$

$$0.42 \leq X \leq 5.62$$

c. Presentase sebaran data dari 99.7% pengamatan

$$99.7\% \rightarrow (\mu - 3\sigma) \leq X \leq (\mu + 3\sigma)$$

$$(3.02 - 3(1.30)) \leq X \leq (3.02 + 3(1.30))$$

$$-0.88 \leq X \leq 6.92$$

### 3.3.8 Dokumentasi dan Penulisan

Tahap terakhir adalah dokumentasi dan penulisan untuk tugas akhir dan artikel ilmiah. artikel ilmiah ditujukan untuk submit pada jurnal terindeks Scopus dan Sinta. Penulisan dilakukan mengikuti ketentuan dan kaidah bahasa sesuai aturan yang berlaku.