

Studi Literatur: Prediksi Kata Berikutnya dengan Metode Recurrent Neural Network

Alwizain Almas Trigresian¹, Nisa Hanum Harani²

¹Program Studi Teknik Informatika, Sekolah Vokasi, Universitas Logistik dan Bisnis Internasional

²Program Studi Teknik Informatika, Sekolah Vokasi, Universitas Logistik dan Bisnis Internasional

¹alwizainalmastrigresian@gmail.com *, ²nisa@ulbi.ac.id

Abstract

Next-word prediction is one of the most frequently used tasks in natural language processing. The Recurrent Neural Network (RNN) method is one method that has been proven to be effective in predicting the next word in a sentence, as it is capable of processing text data with order and context. In this research, various algorithms used in the development of next word prediction using the RNN method were analyzed. Some of these algorithms include LSTM (Long Short-Term Memory) and bidirectional LSTM. The results of this research show that the use of the RNN method in predicting the next word is able to provide better results compared to other methods. However, there are still some challenges that need to be overcome in developing the RNN model to predict the next word. Therefore, further research needs to be done in overcoming these challenges so that the use of the RNN method can be further optimized in predicting the next word in a sentence.

Keywords: Word Prediction, Natural Language Processing, RNN (Recurrent Neural Network), LSTM (Long Short Term Memory), Bidirectional LSTM.

Abstrak

Prediksi kata berikutnya menjadi salah satu tugas yang cukup sering digunakan dalam pemrosesan bahasa alami. Metode RNN (Recurrent Neural Network) menjadi salah satu metode yang terbukti secara efektif dalam memprediksi kata berikutnya dalam sebuah kalimat, karena metode tersebut mampu memproses data teks dengan urutan dan konteks. Pada penelitian ini, dilakukan analisis terhadap berbagai algoritma yang digunakan dalam pengembangan prediksi kata berikutnya menggunakan metode RNN. Beberapa algoritma tersebut antara lain adalah LSTM (Long Short-Term Memory) dan bidirectional LSTM. Hasil dari penelitian ini menunjukkan bahwa penggunaan metode RNN dalam memprediksi kata berikutnya mampu memberikan hasil yang lebih baik dibandingkan dengan metode-metode lainnya. Namun, tentunya masih terdapat beberapa tantangan yang perlu diatasi dalam pengembangan model RNN untuk memprediksi kata berikutnya. Oleh karena itu, perlu dilakukan penelitian lebih lanjut dalam mengatasi tantangan-tantangan tersebut agar penggunaan metode RNN dapat semakin dioptimalkan dalam memprediksi kata berikutnya dalam sebuah kalimat.

Kata kunci: Prediksi Kata, Pemrosesan Bahasa Alami, RNN (Recurrent Neural Network), LSTM (Long Short Term Memory), Bidirectional LSTM.

1. Pendahuluan

Penggunaan kecerdasan buatan atau *artificial intelligence* (AI) dalam kehidupan sehari – hari sudah menjadi hal yang biasa di era modern ini. *Natural Language Processing* (NLP) merupakan salah satu

bagian penting dari AI yang berkontribusi dalam menemukan pendekatan produktif untuk berbicara dengan manusia dan mendapatkan keuntungan dari asosiasi yang terbentuk[1]. NLP sebagai AI dapat digunakan untuk memprediksi kata berikutnya saat mengetik dalam sebuah aplikasi. Prediksi

memperkirakan apa yang akan terjadi pada masa mendatang berdasarkan data yang relevan pada masa lalu secara sistematis dan pragmatis, sehingga diharapkan mampu memiliki objektivitas yang besar [2]. Prediksi kata berikutnya atau dapat disebut sebagai *language modeling* merupakan tugas dasar dari NLP[3]. Terdapat beberapa metode yang dapat digunakan dalam pengaplikasian NLP untuk memprediksi kata berikutnya, salah satunya adalah *Recurrent Neural Network* (RNN).

RNN merupakan *Artificial Neural Networks* (ANN) yang secara khusus dirancang untuk memproses data yang bersambung (*sequential data*) dan data deret waktu (*time series data*)[4]. RNN dapat menyimpan memori (*feedback loop*) yang memungkinkan untuk mengenali pola data dengan baik, kemudian menggunakannya untuk membuat prediksi yang akurat[5]. Hal ini membuat RNN dinilai efektif dan menjadi pilihan yang sesuai dalam memprediksi kata berikutnya. Dengan penggunaan *deep learning* memungkinkan algoritma *neural network* mampu bekerja sangat baik ketika membangun model prediksi melalui masalah yang kompleks[6]. Hal ini membuat RNN dinilai efektif dan menjadi pilihan yang sesuai dalam memprediksi kata berikutnya. Saat ini, pengembangan serta modifikasi model RNN telah banyak dilakukan dan diteliti. Beberapa bentuk modifikasi dari RNN yang paling umum digunakan adalah *Long Short Term Memory* (LSTM) dan *Bidirectional Long Short Term Memory* (Bi-LSTM). LSTM menawarkan kinerja yang lebih baik dibandingkan dengan RNN biasa. LSTM menjadi sub kelas dari *deep learning* yang merupakan pengembangan dari algoritma RNN yang dapat digunakan dalam membuat suatu prediksi dan klasifikasi [7]. Hal ini dikarenakan pada model LSTM ditambahkan *memory cell* yang membuat informasi dapat disimpan dalam jangka waktu yang lama. Selain itu, LSTM dapat mengatasi masalah *vanishing gradient* yang ada pada RNN biasa [8]. Model LSTM biasa digunakan untuk prediksi serta cocok digunakan pada data teks dan data *time series*[9]. Berbeda dengan LSTM, pada model Bi-LSTM memungkinkan informasi yang dimiliki secara dua arah yakni maju dan mundur dengan menggabungkan dua RNN independen secara bersamaan [10]. Algoritma Bi-LSTM dapat melatih *neural network* dengan menggunakan urutan data input pada masa lalu dan masa depan [11]. Algoritma tersebut memproses data masukan dengan menggunakan dua lapisan yang terhubung [12].

Penelitian oleh Sourabh et. al pada tahun 2021 dimana didapatkan hasil bahwa RNN dengan *long short term memory* (LSTM) mampu memprediksi teks dengan akurasi 56% setelah melalui beberapa kali percobaan[1]. Di Indonesia, pada tahun 2022 dilakukan penelitian serupa dengan data destinasi wisata di Indonesia dan

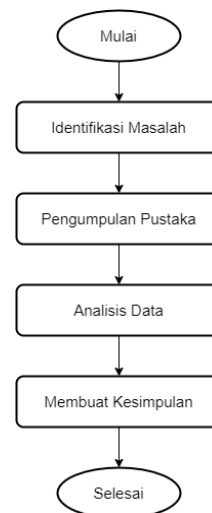
mendapat tingkat akurasi 75% dengan tingkat kesalahan sebesar 55% [13].

Penelitian selanjutnya dilakukan oleh Sharma et. al pada tahun 2019, menghasilkan model *machine learning* prediksi kata berikutnya dengan menggunakan model LSTM dan Bi-LSTM. Pada model LSTM memperoleh akurasi sebesar 70.89%, sedangkan Bi-LSTM memperoleh hasil 79.54% [14]. Penelitian yang dilakukan oleh Ganai A. dan Khursheed F. pada tahun 2019, memperoleh hasil berupa dibuatnya model prediksi kata berikutnya menggunakan algoritma LSTM memperoleh akurasi yang terbaik sebesar 48% [3]. Penelitian lainnya dilakukan oleh Roy et. al pada tahun 2018, menghasilkan aplikasi sederhana untuk solusi kata dan kalimat panjang dalam membantu mempercepat pengetikan. Penelitian ini menggunakan algoritma LSTM yang kemudian mengukur skor *Bits* per karakter (BPC) untuk evaluasinya. Hasil dari evaluasi setelah dilakukan perbandingan dengan kinerja pada penelitian sebelumnya menghasilkan model terbaik (LSTM) dengan skor BPC 1.50 [15].

Berdasarkan hasil dari penelitian terdahulu, pada penelitian ini akan dilakukan studi literatur untuk mengetahui metode yang tepat dalam mengembangkan sistem prediksi kata berikutnya dengan menggunakan salah satu metode RNN. Deep learning memiliki banyak metode untuk prediksi kata berikutnya, namun pada studi literatur ini akan berfokus menggunakan metode RNN dengan algoritma LSTM dan Bi-LSTM.

2. Metode Penelitian

Metodologi penelitian yang digunakan pada penelitian ini menggunakan metode studi literatur yang didapat melalui informasi dari jurnal, buku, atau sumber lain yang telah dibuat sebelumnya [16]. Adapun alur dari penelitian yang dilakukan dalam penelitian ini digambarkan pada Gambar 1.



Gambar 1. Diagram Alur Metodologi Penelitian

2.1. Identifikasi Masalah

Identifikasi masalah digunakan untuk proses pengenalan dan pengidentifikasian suatu masalah yang terjadi saat ini. Masalah yang teridentifikasi berupa beragamnya metode pemrosesan bahasa alami didalam RNN untuk menemukan metode terbaik dalam prediksi kata berikutnya.

Sarkar, Towhid Hussain, Md. Mehedi Hasan, Dewan Md Farid, Swakkhar Shatabda [17]

Somewher ein Blog

2.2. Pengumpulan Pustaka

Pengumpulan pustaka digunakan untuk memperoleh data Pustaka yang relevan dengan masalah yang akan diteliti. Pengumpulan pustaka ini dilakukan sebagai proses pengukuran, pengumpulan, dan validasi data yang akan dibutuhkan dalam penelitian. Adapun macam sumber data yang digunakan pada penelitian ini adalah jurnal, buku, dan sumber lainnya.

Md. Sanzidul Islam, Sadia Sultana Sharmin Mousumi, Sheikh Abujar, Syed Akhter Hossain [4]

Web Scrapping

2019

Bangladesh

2.3. Analisis Data

Pada tahap ini, data yang telah diperoleh akan dilakukan analisis guna memahami metode RNN yang tepat untuk digunakan dalam prediksi kata berikutnya. Tahapan analisis data ini akan membandingkan beberapa metode yang digunakan dalam model prediksi kata berikutnya.

Khrystyna Shakhovska, Iryna Dumyn, Natalia Kryvinska, Mohan Krishna Kagita [18]

Artificial Intelligence Department, Lviv Polytechnic National University

2021

Ukraina

2.4. Membuat Kesimpulan

Tahapan terakhir adalah penarikan kesimpulan yang berdasarkan hasil penelitian dan temuan yang diperoleh dari analisis data yang telah dilakukan.

Aejaz Farooq Ganai, Farida Khursheed [3]

Project Gutenberg

2019

India

3. Hasil dan Pembahasan

Pada penelitian yang dilakukan, memperoleh pustaka penelitian sebelumnya yang terkait dengan informasi seperti pada Tabel 1.

Partha Pratim Barman, Abhijit Boruah [19]

Novel Assam “Bhanumati”

2018

India

Andrew Brusso [20]

Shakespearean and alice in wonderland corpus, facebook, discord, essay

2020

United States

Tabel 1. Pustaka Terkait

Peneliti	Sumber Data	Tahun	Negara
Radhika Sharma, Nishtha Goel, Nishita Aggarwal, Prajyot Kaur, Chandra Prakash [14]	IIT Bombay	2019	India
Afika Rianti, Suprih Widodo, Atikah Dhani, Ayuningtyas, Fadlan Bima Hermawan [13]	Web Scrapping	2022	Indonesia
Sourabh Ambulgekar, Sanket Malewadikar, Raju Garande, Dr. Bharti Joshi [1]	Ebook Author Franz Kafka	2021	India
Sadidul Islam, Mst. Farhana	Protom Alo,	2018	Bangladesh

Shuvendu Roy, Sk. Imran Hossain, M. A. H. Akhand, N. Siddique [15]

Bangla text data Wikipedia, wikidump

2018

Bangladesh

Karma Wangchuk, Tandin Wangchuk and Tenzin Namgye [21]

Dzongkha Development Commission of Bhutan

2023

Bhutan

Dr. S. Rajakumar, Dr. V. Rameshbabu,

Nietzsche Text

2023

India

Dr. D. Usha, Nikila K., Ramya Shree B., Sakthi Priya R. [22]				Namgail, Sandeep Kumar [32]	dictionarie s (Bodhi words)			
K. Chakradhar, K. Sai Kiran, K. Shanmukh, K. Sharath Kumar, K. Dinesh Sagar [23]	Amharic sentence data	2022	India	S. Ramya, C.S. Kanimozhi Selvi [33]	The New York Times news (Kaggle)	2019		India
Demeke Endalie, Getamesay Haile, Wondmagegn Taye [24]	Amharic bible	2022	Euthopia	P. Niharika, S. John Justin Thangaraj [34]	Gutenberg	2023		India
Rutuja Gosavi, Pradnya Konduskar, Pranav Chavan, Jyoti Kundale [25]	Kaggle dataset	2022	India	Deepal S. Thakur, Rajiv N. Tarsarya, Akshay A. Vaskar, Ashwini Save [35]	-	2019		India
Milind Soam, Sanjeev Thakur [26]	-	2022	India	Muhammad Faizan Khalid [36]	Enron dataset, Persona chat dataset	2021		Swedia
A. Srinagesh, Ch. Sudha Sree, B. Prasanthi, P. Rama Krishna, P. Siva Prasad [27]	Twitter API Telugu Language Tweets	2021	India	Ferialle Lahrache, Sana Djebrit [37]	Coursera Swiftkey, Kaggle, Brown Corpus	2020		Algeria
Keerthana N., Harikrishnan S., Konsaha Buji M., Jona J. B [28]	Irish Lyrics (English)	2021	India	Mahir Mahbub, Suravi Akhter, Ahmedul Kabir, Zerina Begum [38]	Bengali Newspaper Articles, Bengali Wikipedia	2022		Bangladesh
Yuerong Ma, Yating Wang, Chengjie Wan, Tian Qiao [29]	Sports and Financial Dataset (BBC Dataset)	2021	China	Ali Pourmohammad, Mensur Gulami, Javid Mahmudov, Yusif Aliyev, Rovshan Akberov, Anar Sultani [39]	Azeri Vocabulary Database	2020		Azerbaijan
Kolambi Narula [30]	-	2023	India	Tiwalayo N. Eisape, Noga Zaslavsky, Roger P. Levy [40]	Provo Corpus	2020		Amerika
Chinmaya Nayak, Arvind Kumar [31]	-	2022	India					
Ankush Kumar, Pushpendra Kumar Mishra, Tsering	Ladakhi articles, newspaper	2023	India					

Setelah memperoleh pustaka atau penelitian sebelumnya yang terkait, selanjutnya dilakukan analisis data dengan memilih pustaka yang paling sesuai untuk memahami metode yang tepat dalam prediksi kata berikutnya. Pustaka yang terpilih disajikan pada Tabel 2.

Tabel 2. Pustaka Terpilih

Topik	Model	Akurasi	BPC	Prediction Using Machine Learning [22]			
Next Words Prediction Using Recurrent Neural Networks [1]	LSTM	56%	-	Next Word Prediction Using Deep Learning [23]	LSTM	75.02%	-
Next Word Prediction Using LSTM [13]	LSTM	75%	-	Bi-Directional Long Short Term Memory-Gated Recurrent Unit Model for Amharic Next Word Prediction [24]	Bi- LSTM	76.1%	-
Next Word Prediction in Hindi Using Deep Learning Techniques [14]	Bi-LSTM	79.54%	-	Next Word Prediction Using Recurrent Neural Network [25]	LSTM	58.6%	-
Predicting next Word using RNN and LSTM cells: Stastical Language Modeling [3]	LSTM	48%	-	Next Word Prediction Using Deep Learning: A Comparative Study [26]	Bi-LSTM	93%	-
Sequence Modeling for Intelligent Typing Assistant with Bangla and English Keyboard [15]	LSTM	-	1.50	Next Word Prediction In Telugu Sentences Using Recurrent Neural Networks [27]	LSTM	70%	-
An Approach for a Next-Word Prediction for Ukrainian Language [18]	LSTM, Markov Chains	74.55%	-	Next Word Prediction Using Machine Learning Techniques [31]	LSTM	80%	-
A RNN based Approach for next word prediction in Assamese Phonetic Transcription [19]	LSTM	72.10%	-	Next Word Prediction in Bodhi Language Using LSTM based Approach [32]	LSTM	92-96%	-
Dzongkha Next Words Prediction Using Bidirectional LSTM [21]	Bi-LSTM	73.89%	-	Recurrent Neural Network based Models for Word Prediction [33]	RNN LSTM	62% 64%	- -
EXO Next Word	Bi-LSTM	81.07%	-	Long Short Term Memory	Bi-LSTM	72%	-
					LSTM	48%	-

model-based automatic next word generation for text-based applications In contrast to the N-gram model [34]				
Text Prediction Using Machine Learning [36]	LSTM	80.69%	-	
	GRU	79%	-	
Next Word Prediction Based On Deep Learning [37]	RNN	71.51%	-	

Berdasarkan analisis data pada Tabel 2 diperoleh hasil prediksi kata berikutnya dengan akurasi sebesar 56% dengan menggunakan model LSTM pada topik “*Next Words Prediction Using Recurrent Neural Network*” [1]. Penelitian tersebut juga telah melewati beberapa percobaan dan memperoleh hasil akurasi terbaiknya. Selanjutnya pada penelitian dengan topik “*Next Word Prediction Using LSTM*” memperoleh model LSTM untuk membangun prediksi kata berikutnya dengan akurasi sebesar 75% dengan tingkat *loss* atau kesalahan sebesar 55% [13]. Penelitian ini memprediksi kata berikutnya dari suatu destinasi wisata di Indonesia. Hasil akurasi yang diperoleh menunjukkan lebih baik daripada penelitian sebelumnya yang menggunakan model berbeda.

Pada penelitian dengan topik “*Next Word Prediction in Hindi Using Deep Learning Techniques*” diperoleh model *machine learning* prediksi kata berikutnya dalam bahasa Hindi dengan menggunakan model LSTM dan Bi-LSTM. Pada model LSTM memperoleh akurasi sebesar 70.89%, sedangkan Bi-LSTM memperoleh hasil 79.54%. Bi-LSTM memiliki hasil yang lebih baik jika dibandingkan dengan LSTM [14]. Kemudian pada penelitian dengan topik “*Predicting next Word using RNN and LSTM cells: Stastical Language Modeling*” memperoleh hasil berupa dibuatnya model prediksi kata berikutnya dengan menggunakan teknik *many-to-one model* dan *many-to-many model* [3]. Data yang digunakan dari kumpulan buku *project Gutenberg*. Dari model yang telah dibuat dilakukan perbandingan dengan hasil *many-to-one model* memperoleh akurasi yang terbaik sebesar 48%, sedangkan untuk *many-to-many model* memperoleh akurasi sebesar 19%.

Penelitian dengan topik “*Sequence Modeling for Intelligent Typing Assistant with Bangla and English Keyboard*” menghasilkan aplikasi sederhana untuk solusi kata dan kalimat panjang dalam membantu mempercepat pengetikan [15]. Penelitian ini

menggunakan algoritma LSTM dengan menggunakan 2 metode yaitu *generating single long sequence* yang berbasis *encoder-decoder* dengan mengambil teks yang telah diberikan dan memahami konteks kalimat tersebut untuk dihasilkan urutan kata berikutnya. Kemudian metode berikutnya adalah *beam search* yang menyarankan beberapa kata berdasarkan konteks terdekat. Model pada penelitian ini bekerja untuk melatih 2 bahasa, yaitu bahasa Inggris dan Bangla. Kemudian mengukur skor *Bits per karakter (BPC)* untuk evaluasinya. Hasil dari evaluasi setelah dilakukan perbandingan dengan kinerja pada penelitian sebelumnya menghasilkan model terbaik (LSTM) dengan skor BPC 1.50 (lebih rendah daripada skor BPC penelitian sebelumnya). Sehingga model LSTM digunakan untuk membuat model saran kata bahasa Bangla. Penelitian dengan topik “*An Approach for a Next-Word Prediction for Ukrainian Language*” memperoleh hasil algoritma LSTM dan *Markov Chains* untuk prediksi kata dalam bahasa Ukraina dengan hasil akurasi sebesar 74.55% [18]. Penelitian dengan topik “*A RNN based Approach for next word prediction in Assamese Phonetic Transcription*” menghasilkan model RNN untuk prediksi kata berikutnya dengan bahasa Assamese menggunakan algoritma LSTM. Model tersebut memperoleh akurasi sebesar 72.10% sehingga model tersebut dapat digunakan dalam memprediksi kata berikutnya dari bahasa Assamese pada saat pengetikan fonetik [19].

Penelitian selanjutnya dengan topik “*Dzongkha Next Words Prediction Using Bidirectional LSTM*” menghasilkan sistem prediksi kata berikutnya pada suku kata Dzongkha dengan menggunakan model Bi-LSTM yang dapat mencapai akurasi terbaik sebesar 73.89% [21]. Penelitian lainnya dengan topik “*EXO Next Word Prediction Using Machine Learning*” menghasilkan sistem prediksi kata berikutnya dengan algoritma Bi-LSTM yang menggunakan *input* data bahasa Inggris memperoleh hasil akurasi sebesar 81.07% [22]. Penelitian dengan topik “*Next Word Prediction Using Deep Learning*” dan “*Bi-Directional Long Short Term Memory-Gated Recurrent Unit Model for Amharic Next Word Prediction*” menggunakan model LSTM serta Bi-LSTM dalam memprediksi kata berikutnya pada suku kata Amharic [23] [24]. Pada kedua penelitian tersebut didapatkan hasil akurasi yang sama sebesar 75.02% untuk LSTM dan 76.1% untuk Bi-LSTM. Penelitian serupa dilakukan dengan sumber data Kaggle dataset, penelitian dengan topik “*Next Word Prediction Using Recurrent Neural Network*” mendapatkan hasil akurasi yang sangat berbeda antara LSTM dan Bi-LSTM yakni sebesar 58.6% dan 93% [25].

Penelitian dengan topik “*Next Word Prediction Using Deep Learning: A Comparative Study*” menghasilkan model prediksi kata berikutnya dengan pemanfaatan algoritma *deep learning* LSTM dan Bi-LSTM dengan

perolehan akurasi sebesar 66.1% untuk Bi-LSTM dan 58.27% untuk LSTM [26]. Penelitian berikutnya dengan topik “*Next Word Prediction In Telugu Sentences Using Recurrent Neural Networks*” memperoleh model *machine learning* yang dapat digunakan untuk prediksi kata berikutnya menggunakan bahasa Telugu dengan algoritma LSTM [27]. Model tersebut memperoleh akurasi terbaiknya sebesar 70%. Kemudian pada penelitian dengan topik “*Next Word Prediction*” dan “*Next Word Prediction Using Machine Learning Techniques*” diperoleh model prediksi kata berikutnya dengan menggunakan algoritma LSTM yang menghasilkan akurasi sebesar 80% dan 92% hingga 96% [28] [31]. Penelitian dengan topik “*Next Word Prediction in Bodhi Language Using LSTM based Approach*” memperoleh model *machine learning* berupa LSTM untuk memprediksi kata berikutnya dengan menggunakan bahasa Ladakhi atau Bodhi [32]. Model tersebut memperoleh akurasi sebesar 50%.

Penelitian lainnya dengan topik “*Recurrent Neural Network based Models for Word Prediction*” menggunakan tiga model RNN dalam membuat prediksi kata berikutnya menggunakan sumber data bahasa Inggris, yaitu RNN, LSTM, dan Bi-LSTM [33]. Masing-masing model memperoleh akurasi sebesar 62%, 64%, dan 72%. Model Bi-LSTM menjadi model yang terbaik dengan *dataset* yang digunakan. Pada penelitian dengan topik “*Long Short Term Memory model-based automatic next word generation for text-based applications In contrast to the N-gram model*” menggunakan algoritma LSTM untuk memprediksi kata berikutnya dengan data diambil dari buku pada *website* Gutenberg memperoleh hasil akurasi sebesar 48% [34]. Akurasi tersebut lebih baik jika dibandingkan dengan model N-gram. Penelitian dengan topik “*Text Prediction Using Machine Learning*” menghasilkan model prediksi kata berikutnya dari data percakapan dengan menggunakan algoritma LSTM dan GRU yang memiliki akurasi masing-masing sebesar 80.69% dan 79% [36]. LSTM menjadi algoritma dengan akurasi terbaik pada penelitian tersebut. Penelitian selanjutnya dengan topik “*Next Word Prediction Based On Deep Learning*” menyebutkan bahwa tingkat akurasi RNN dalam memprediksi kata berikutnya dinilai sebesar 71.51% [37]. Model RNN tersebut telah dilakukan percobaan pada tiga *dataset* berbeda dan memperoleh akurasi terbaiknya sebesar 71.51%.

Melalui analisis data yang diperoleh tersebut, algoritma LSTM memiliki kemampuan untuk mempelajari dan mengingat informasi jangka panjang pada urutan data masukan. Penggunaan algoritma LSTM akan lebih baik untuk memprediksi kata berikutnya pada huruf alfabet internasional. Sedangkan disisi lain, Bi-LSTM adalah jenis arsitektur RNN yang memproses urutan data masukan dengan menggabungkan dua arah dari masukan sebelumnya dan selanjutnya. Bi-LSTM dapat

menggabungkan informasi dari kedua arah dan memperbaiki kelemahan dalam mengingat informasi jangka panjang pada arsitektur LSTM. Penggunaan algoritma Bi-LSTM mampu menjadi solusi dalam prediksi kata berikutnya dengan huruf selain alfabet internasional, seperti suku kata Dzongkha, Bahasa Hindi, bahasa Amharic, bahasa Bodhi, dan bahasa Telugu.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa prediksi kata berikutnya dengan menggunakan metode RNN, khususnya LSTM dan Bi-LSTM telah banyak dipelajari dalam beberapa penelitian terbaru. Hasil penelitian menunjukkan bahwa penggunaan model LSTM dan Bi-LSTM dapat menghasilkan akurasi yang cukup baik dalam memprediksi kata berikutnya pada berbagai bahasa dan topik, seperti bahasa Inggris, bahasa Hindi, bahasa Ukraina, bahasa Assamese, suku kata Dzongkha, bahasa Indonesia, bahasa Amharic, bahasa Bodhi, dan bahasa Telugu. Namun penggunaan LSTM akan lebih baik untuk memproses bahasa dengan struktur alfabet internasional, sedangkan penggunaan Bi-LSTM lebih baik digunakan untuk memproses bahasa dengan struktur selain alfabet internasional, seperti suku kata Dzongkha, bahasa Hindi, bahasa Amharic, bahasa Bodhi, dan bahasa Telugu. Namun, penggunaan algoritma Bi-LSTM juga dapat digunakan untuk memproses bahasa dengan struktur alfabet internasional. Meskipun demikian, penelitian lebih lanjut sangat diperlukan untuk mengetahui pengaplikasian lainnya terkait prediksi kata berikutnya dengan metode RNN, terkhusus LSTM dan Bi-LSTM.

Daftar Rujukan

- [1] S. Ambulgekar, S. Malewadikar, R. Garande, and B. Joshi, “Next Words Prediction Using Recurrent Neural Networks,” *ITM Web of Conferences*, vol. 40, p. 03034, Aug. 2021, doi: 10.1051/itmconf/20214003034.
- [2] I. Admirani, “Penerapan Metode Fuzzy Time Series Untuk Prediksi Laba Pada Perusahaan,” *Jurnal JUPITER*, vol. 10, no. 1, pp. 19–31, 2018.
- [3] A. F. Ganai and F. Khursheed, “Predicting next Word using RNN and LSTM cells: Stastical Language Modeling,” 2019.
- [4] I. Md. Sanzidul, S. M. Sadiq Sultana, S. Abujar, and S. A. Hossain, “Sequence-to-sequence Bangla sentence generation with LSTM recurrent neural networks,” in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 51–58. doi: 10.1016/j.procs.2019.05.026.
- [5] J. Wira and G. Putra, *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*. 2020.
- [6] B. Ekaba, *Building Machine Learning and Deep Learning Models on Google Cloud Platform (A Comprehensive Guide for Beginners-Apress)*. 2019. doi: 9781484244708.
- [7] R. Mugi Setya Adi, “Prediksi Harga Komoditas Pangan Menggunakan Algoritma Long Short-Term Memory (LSTM),” *Building of Informatics Technology and Science (BITS)*, vol. 4, no. 2, 2022, doi: 10.47065/bits.v4i2.2229.
- [8] N. K. Manaswi, *Deep Learning with Applications Using Python*. Berkeley, CA: Apress, 2018. doi: 10.1007/978-1-4842-3516-4.

- [9] R. Dwi, W. Santosa, M. Arif Bijaksana, and A. Romadhony, "Implementasi Algoritma Long Short-Term Memory (LSTM) untuk Mendeteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia," *Jurnal Tugas Akhir Fakultas Informatika*, vol. 8, no. 1, p. 691, 2021.
- [10] J. Xie, B. Chen, X. Gu, F. Liang, and X. Xu, "Self-Attention-Based BiLSTM Model for Short Text Fine-Grained Sentiment Classification," *IEEE Access*, vol. 7, pp. 180558–180570, 2019, doi: 10.1109/ACCESS.2019.2957510.
- [11] M. J. Hamayel and A. Y. Owda, "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms," *AI*, vol. 2, no. 4, pp. 477–496, Oct. 2021, doi: 10.3390/ai2040030.
- [12] K. A. Althelaya, E.-S. M. El-Alfy, and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," in *2018 9th International Conference on Information and Communication Systems (ICICS)*, IEEE, Apr. 2018, pp. 151–156. doi: 10.1109/IACS.2018.8355458.
- [13] A. Rianti, S. Widodo, A. D. Ayuningtyas, and F. B. Hermawan, "Next Word Prediction Using LSTM," *Journal of Information Technology and Its Utilization*, vol. 5, no. 1, 2022.
- [14] R. Sharma, N. Gael, N. Aggarwal, and C. Prakash, "Next Word Prediction in Hindi Using Deep Learning Techniques," 2019.
- [15] S. Roy, S. Imran Hossain, M. A. H Akhand, and N. Siddique, "Sequence Modeling for Intelligent Typing Assistant with Bangla and English Keyboard," 2018.
- [16] M. S. Rumetna, "Pemanfaatan Cloud Computing pada Dunia Bisnis: Studi Literatur," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 3, p. 305, Aug. 2018, doi: 10.25126/jtiik.201853595.
- [17] S. Islam, Mst. F. Sarkar, T. Hussain, M. M. Hasan, D. M. Farid, and S. Shatabda, "Bangla Sentence Correction Using Deep NeuralNetwork Based Sequence to Sequence Learning," *International Conference on Computer and Information Technology (ICCIT)*, 2018.
- [18] K. Shakhovska, I. Dumyn, N. Kryvinska, and M. K. Kagita, "An Approach for a Next-Word Prediction for Ukrainian Language," *Wirel Commun Mob Comput*, vol. 2021, 2021, doi: 10.1155/2021/5886119.
- [19] P. P. Barman and A. Boruah, "A RNN based approach for next word prediction in assamese phonetic transcription," in *Procedia Computer Science*, Elsevier B.V., 2018, pp. 117–123. doi: 10.1016/j.procs.2018.10.359.
- [20] A. Brusso, "Text Generation with Recurrent Neural Networks," 2020.
- [21] K. Wangchuk, T. Wangchuk, and T. Namgyel, "Dzongkha Next Words Prediction Using Bidirectional LSTM," *Bhutan Journal of Research and Development*, no. 2, Feb. 2023, doi: 10.17102/bjrd.rub.se2.038.
- [22] S. Rajakumar, V. Rameshbabu, D. Usha, N. K. R. Shree, and S. Priya, "EXO Next Word Prediction Using Machine Learning," *J Surv Fish Sci*, vol. 10, no. 3S, pp. 4112–4118, 2023.
- [23] K. Chakradhar, K. S. Kiran, K. Shanmukh, K. Sharath Kumar, K. Dinesh Sagar, and] Gmr, "Next Word Prediction Using Deep Learning," 2022. [Online]. Available: www.ijrpr.com
- [24] D. Endalie, G. Haile, and W. Taye, "Bi-directional long short term memory-gated recurrent unit model for Amharic next word prediction," *PLoS One*, vol. 17, no. 8 August, Aug. 2022, doi: 10.1371/journal.pone.0273156.
- [25] R. Gosavi, P. Konduskar, P. Chavan, and J. Kundale, "Next Word Prediction Using Recurrent Neural Network," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 9, no. 8, 2022, [Online]. Available: www.jetir.org
- [26] M. Soam and S. Thakur, "Next Word Prediction Using Deep Learning: A Comparative Study," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, IEEE, Jan. 2022, pp. 653–658. doi: 10.1109/Confluence52989.2022.9734151.
- [27] A. Srinagesh, C. Sudha Sree, B. Prasanthi, M. P. Rama Krishna, and M. P. S. Prasad, "Next Word Prediction in Telugu Sentences Using Recurrent Neural Networks," *Turkish Online Journal of Qualitative Inquiry (TOJQI)*, vol. 12, no. 5, pp. 4284–4292, 2021.
- [28] Keerthana, Harikrishnan, Konsaha Buji, and Jona J, "Next Word Prediction," *International Journal of Creative Research Thoughts*, vol. 9, no. 12, 2021, [Online]. Available: www.ijcrt.org
- [29] Y. Ma, Y. Wang, C. Wan, and T. Qiao, "Design of Word Input Prediction System Based on LSTM," *International Journal of Scientific Research and Innovative Technology*, vol. 8, no. 8, 2021, [Online]. Available: www.ijrsit.com
- [30] K. Narula, "A Critical Review on Next Word Prediction," *International Journal of Advanced Research in Science, Communication and Technology (IJARST)*, vol. 3, no. 1, 2023, doi: 10.48175/568.
- [31] C. Nayak and A. Kumar, "Next Word Prediction Using Machine Learning Techniques," *Advanced Engineering Science*, vol. 54, no. 2, 2022, [Online]. Available: https://www.researchgate.net/publication/365870747
- [32] A. Kumar, P. K. Mishra, T. Namgail, and S. Kumar, "Next Word Prediction in Bodhi Language Using LSTM based Approach," 2023, [Online]. Available: https://ssrn.com/abstract=4367666
- [33] Ms. S. Ramya and Dr. C. S. K. Selvi, "Recurrent Neural Network based Models for Word Prediction," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 4, pp. 7433–7437, Nov. 2019, doi: 10.35940/ijrte.D5313.118419.
- [34] P. Niharika and S. J. J. Thangaraj, "Long Short Term Memory model-based automatic next word generation for text-based applications In contrast to the N-gram model," *J Surv Fish Sci*, vol. 10, no. 1S, 2023.
- [35] D. S. Thakur, R. N. Tarsarya, A. A. Vaskar, and A. Save, "A Survey on Text Prediction Techniques," *International Journal for Research and Innovation*, vol. 1, no. 2, 2019, [Online]. Available: www.viva-technology.org/New/IJRI
- [36] M. F. Khalid, "Text Prediction Using Machine Learning," Linköping University, Sweden, 2021.
- [37] F. Lahrache and S. Djebrit, "Next Word Prediction Based On Deep Learning," University of Ghardaia, Algeria, 2020.
- [38] M. Mahbub, S. Akhter, A. Kabir, and Z. Begum, "Context-based Bengali Next Word Prediction: A Comparative Study of Different Embedding Methods," *Dhaka University Journal of Applied Science and Engineering*, vol. 7, no. 2, pp. 8–15, Apr. 2023, doi: 10.3329/dujase.v7i2.65088.
- [39] A. Pourmohammad, M. Gulami, J. Mahmudov, Y. Aliyev, R. Akberov, and A. Sultani, "The First Azeri (Azerbaijani) Language Next Word Predictor," *Information Systems and Signal Processing Journal*, vol. 5, no. 1, 2020, doi: 10.23977/isspj.2020.51001.
- [40] T. N. Eisape, N. Zaslavsky, and R. P. Levy, "Cloze Distillation: Improving Neural Language Models with Human Next-Word Predictions," in *Proceedings of the 24th Conference on Computational Natural Language Learning*, Association for Computational Linguistics (ACL), 2020. doi: 10.18653/v1/P17.

