

Next Word Prediction pada Rekam Medis Elektronik menggunakan Gated Recurrent Unit

Dian Prawira ^[1]; Ilhamsyah ^[2]

Program Studi Sistem Informasi, Fakultas MIPA^[1] ^[2]
Universitas Tanjungpura Pontianak
dianprawira@sisfo.untan.ac.id

INFO ARTIKEL	INTISARI
Diajukan : 05 Oktober 2024	<i>Tujuan utama dari penelitian ini adalah untuk membuat model Gated Recurrent Unit (GRU) yang dirancang khusus untuk meramalkan kata berikutnya dalam rekam medis elektronik (RME). Pembangunan model ini melibatkan serangkaian proses, termasuk membangun struktur dan parameter serta melakukan pelatihan menggunakan dataset dengan 169.544 token. Data pelatihan dibagi dengan rasio 80% untuk pelatihan dan 20% untuk pengujian. Evaluasi kinerja dilakukan selama 90 iterasi, menghasilkan penurunan loss function dari 2,2836 menjadi 0,800 dan peningkatan akurasi dari 60,26% menjadi 80,59%. Model GRU mencapai akurasi prediksi 87,04% pada dataset pengujian. Penelitian ini mengimplikasikan pemanfaatan teknologi deep learning untuk mengoptimalkan proses administratif di fasilitas kesehatan dan memberikan kontribusi dalam pengembangan sistem prediktif untuk aplikasi medis lainnya.</i>
Diterima : 22 November 2024	
Diterbitkan: 19 Desember 2024	
Kata Kunci : <i>Gated Recurrent Unit, Next Word Prediction, Rekam Medik Elektronik, Natural Language Processing</i>	

I. PENDAHULUAN

Rekam medis elektronik (RME) adalah salah satu komponen penting dalam sektor kesehatan modern. RME menyediakan catatan digital yang komprehensif tentang riwayat kesehatan pasien, termasuk diagnosis, pengobatan, dan hasil laboratorium. Penggunaan RME dapat meningkatkan kualitas perawatan pasien melalui akses cepat dan mudah ke informasi medis yang akurat dan terkini (Boonstra, Versluis, & Vos, 2014).

Meskipun RME menawarkan banyak keuntungan, mengelolanya pasti menantang. Salah satu masalah utama adalah kompleksitas data yang akan diproses dan disimpan. Data medis seringkali tidak terstruktur dan memiliki variasi yang sangat besar, mulai dari teks bebas hingga gambar medis. Sekitar 80% data medis adalah data yang tidak terstruktur sehingga membuat proses analisis dan manajemen menjadi lebih rumit. Selain itu, masalah keamanan dan privasi data pasien juga menjadi perhatian utama dalam pengelolaan RME (D. Zhang, Yin, Zeng, Yuan, & Zhang, 2020).

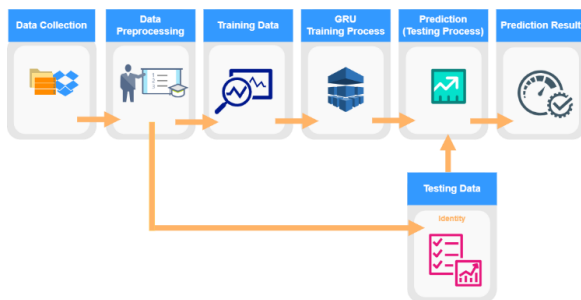
Teknologi *Artificial Intelligence* (AI) dan *Machine Learning* (ML) telah menunjukkan potensi besar dalam memfasilitasi pengelolaan data medis. Salah satu aplikasi AI yang paling menjanjikan adalah *Natural Language Processing* (NLP), yang memungkinkan komputer memahami dan memproses teks manusia (Harrison & Sidey-Gibbons, 2021). NLP dapat digunakan untuk mengekstrak informasi penting dari rekam medis,

mengurangi beban kerja tenaga medis, dan meningkatkan akurasi diagnosis (Sidey-Gibbons & Sidey-Gibbons, 2019).

Next Word Prediction (Prediksi Kata Berikutnya) dalam teks medis adalah salah satu aplikasi NLP yang dapat meningkatkan efisiensi dan akurasi pengisian rekam medis. Model prediksi kata dapat membantu tenaga medis dalam mengetik rekam medis lebih cepat dan mengurangi kesalahan pengetikan. Salah satu teknik yang digunakan untuk prediksi kata adalah *Gated Recurrent Unit* (GRU) yang merupakan variasi dari *Recurrent Neural Network* (RNN) (Arfianti, Novitasari, Widodo, Hafiyusholeh, & Utami, 2021; Krishnan, Magalingam, & Ibrahim, 2021; Pierre, Akim, Semenyo, & Babiga, 2023). GRU memiliki keunggulan dalam menangani masalah *vanishing gradient* yang umum terjadi pada RNN tradisional, sehingga lebih efektif dalam memproses data teks panjang (Chung, Gulcehre, Cho, & Bengio, 2014).

II. BAHAN DAN METODE

Metode penelitian ini dimulai dari tinjauan pustaka dengan mencari referensi teoritis dari penelitian sebelumnya terkait topik penelitian, yaitu penelitian algoritma GRU untuk memprediksi kata berikutnya dalam sistem rekam medis.



Gambar 1 Teknik Penelitian Prediksi Kata Berikutnya Menggunakan Algoritma GRU

Selanjutnya pada Gambar 1 menunjukkan alur kerja proses prediksi kata berikutnya menggunakan model GRU. Mulai dari pendataan yaitu data rekam medis yang diperoleh di salah satu fasilitas kesehatan yang ada di Provinsi Kalimantan Barat dan tambahan *datasheet* yang diambil di Kaggle. Kumpulan data Kaggle yang digunakan dalam penelitian ini secara khusus untuk melengkapi data rekam medis, untuk memastikan analisis yang komprehensif.

Kemudian, data yang terkumpul diolah sebelum digunakan sebagai data pelatihan. Setelah itu, data dilatih menggunakan proses pelatihan GRU. Sebagian data digunakan untuk menguji data. Model yang dilatih kemudian digunakan untuk prediksi dalam proses pengujian, yang menghasilkan hasil prediksi. Selain itu, ada data pengujian yang digunakan untuk menguji model, di mana hasil pengujian juga berkontribusi pada proses prediksi.

Pengujian model dalam kasus prediksi kata selanjutnya akan menghasilkan probabilitas atau nilai kontinu untuk setiap kata, oleh karena itu relevan digunakan pengujian dengan pengukuran *Mean Square Error* (MSE) (Y. Zhang & Yang, 2022). Namun, jika model menghasilkan kata atau token sebagai output, evaluasi dengan metrik diskrit seperti akurasi lebih tepat.

III. HASIL DAN PEMBAHASAN

A. Dataset Description

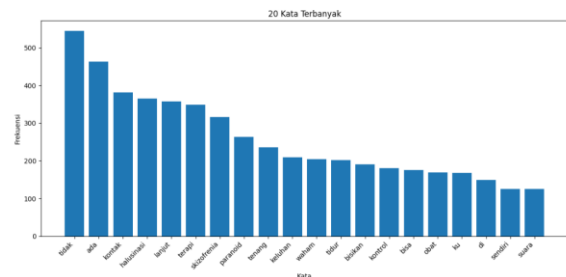
Dataset yang digunakan dalam penelitian ini memiliki atribut sebagai berikut: Id Pasien, Tanggal, Subjektif, Tujuan, Penilaian, Rencana. Atribut Gabungan adalah kombinasi dari Subjektif, Tujuan, Penilaian, dan Rencana. Gambar 2 adalah data rekam medis elektronik yang terdiri dari rekam medis dalam bahasa Indonesia.

Id Pasien	tanggal	Subjektif	Objektif	Assesment	Plan	Sabungan
136513.0	2020-07-09 00:00:00	kantak, idur, bayang, biang	Kontak (+) ku tenang, E/A stabil/berbatas, wah.	skizofrenia residual	terapi lanjut	kantak, idur, bayang, biang Kontak (+) ku te.
NaH	2020-08-26 00:00:00	Kontrol, idur bisa, suara biskak, bayangan...	Kontak (+) ku tenang, E/A stabil/berbatas, wah.	skizofrenia residual	terapi lanjut	Kontrol, idur bisa, suara biskak, bayangan...
NaH	2020-10-06 00:00:00	kontrol, kabar baik, tidak ada keluhan, obat c...	Kontak (+) ku tenang, E/A terbatas, wah, m...	skizofrenia residual	terapi lanjut	kontrol, kabar baik, tidak ada keluhan, obat c...
NaH	2020-11-17 00:00:00	kontrol, tidak ada keluhan, obat cocok, bisa m...	Kontak (+) ku tenang, E/A stabil/berbatas, wah.	skizofrenia residual	terapi lanjut	kontrol, tidak ada keluhan, obat cocok, bisa m...
NaH	2021-01-04 00:00:00	keluhan tidak ada, obat sudah habis	Kontak (+) ku tenang, E/A stabil/berbatas	skizofrenia residual	terapi lanjut	keluhan tidak ada, obat sudah habis

Gambar 2 Dataset Rekam Medik Elektronik

B. Data Preprocessing

Pengolahan data dalam penelitian ini meliputi beberapa langkah, yaitu: (1) pembersihan data, (2) tokenisasi teks, dan (3) normalisasi teks. Pembersihan data dilakukan untuk menghapus data yang tidak relevan atau mengandung kesalahan. Tokenisasi teks dilakukan untuk memecah teks menjadi unit yang lebih kecil, seperti kata atau frasa. Normalisasi teks dilakukan untuk mengubah teks menjadi bentuk yang konsisten, seperti mengubah semua huruf menjadi huruf kecil.



Gambar 3 Sebagian distribusi kata tertinggi pada dataset

Berdasarkan analisis data yang disajikan pada Gambar 3, terbukti bahwa kata "tidak" adalah yang paling dominan, muncul lebih dari 500 kali. Ini menunjukkan bahwa negasi sering digunakan dalam kumpulan data ini, yang berisi total 169.544 token. Ini menunjukkan bahwa negasi sering digunakan dalam dataset ini, dan model prediksi kata berikutnya harus mampu mengenali pola negasi ini dan memprediksi kata-kata yang logis dan relevan dalam konteks negasi, seperti "tidak ada" atau "tidak mungkin". Selain itu, kata-kata seperti "halusinasi", "skizofrenia", "terapi", dan "obat" menunjukkan bahwa kumpulan data terkait erat dengan kesehatan mental, yang berarti model prediksi perlu dilatih untuk memahami pola teks medis dan psikologis di mana istilah-istilah ini umum digunakan.

Penelitian ini menggunakan tokenisasi untuk membagi kata utama dari kata berikutnya dan menggunakan metode GRU untuk mengamati urutan kata dalam arah mundur dan maju. Untuk tujuan ini, digunakan bahasa pemrograman Python. Hasil implementasi dan *preprocessing* yang diperoleh dari penelitian ini dapat dilihat pada Gambar 4.

Learning rate adalah salah satu parameter penting yang mempengaruhi kecepatan dan kualitas pelatihan. Dengan berbagai ukuran yang berbeda-beda, penentuan learning rate yang optimal sangat penting untuk memastikan model dapat belajar dengan efektif. Jika learning rate terlalu tinggi, model akan mengalami osilasi atau bahkan divergensi. Sebaliknya, jika learning rate terlalu rendah, proses pelatihan akan memakan waktu yang sangat lama. Oleh karena itu, penentuan learning rate yang optimal sangat penting untuk memastikan model dapat belajar dengan efektif. Ada beberapa metode untuk menentukan learning rate yang optimal, seperti menggunakan metode grid search, random search, atau menggunakan metode yang lebih canggih seperti AdamW. Dalam penelitian ini, digunakan metode grid search untuk menentukan learning rate yang optimal. Hasilnya menunjukkan bahwa learning rate yang optimal adalah 0.001. Dengan menggunakan learning rate ini, model dapat mencapai akurasi yang tinggi dan stabil.

Gambar 4 Output Token Rekam Medis Elektronik

C. Training Data

Pengembangan model GRU dalam penelitian ini meliputi beberapa langkah, yaitu: (1) penentuan arsitektur model, (2) menentukan parameter dan (3) melatih model menggunakan data rekam medis. Arsitektur model GRU yang digunakan dalam penelitian ini terdiri dari beberapa lapisan GRU yang dihubungkan secara berurutan. Tabel 1 berisi parameter yang digunakan dan disesuaikan berdasarkan hasil eksperimen awal.

Tabel 1 Output Token Rekam Medis Elektronik

Layer (type)	Value
Embedding	64
Learning Rate	0.001
GRU	150
Dropout	0.2
GRU2	100
Dense	2901
Training Data Ratio	80 %
Testing Data Ratio	20 %

Model yang ditampilkan memiliki beberapa lapisan, termasuk Embedding, dua lapisan GRU, Dropout, dan Dense. Lapisan Embedding mengubah input teks menjadi vektor 64 dimensi, diikuti oleh dua lapisan GRU yang menangani urutan data dengan masing-masing 150 dan 100 unit. Lapisan Dropout digunakan untuk mencegah overfitting. Penggunaan dua lapisan GRU dengan dropout 0,2 membantu mengurangi overfitting dan meningkatkan generalisasi model sementara lapisan Dense di akhir bertanggung jawab atas prediksi akhir dengan 2901 unit output.

Selanjutnya, pemilihan model terbaik dilakukan untuk menentukan model dengan kinerja terbaik. Berdasarkan Tabel 2, selama pelatihan skenario 3 pola iterasi, ada peningkatan akurasi secara bertahap dan penurunan *loss function*. Gambar 5 menunjukkan grafik akurasi model selama proses pelatihan yang diukur selama beberapa iterasi. Sumbu X mewakili jumlah iterasi, di mana seluruh himpunan data digunakan untuk memperbarui bobot model di setiap iterasi. Sementara itu, sumbu Y menunjukkan akurasi model, yang merupakan persentase prediksi yang benar dibandingkan dengan prediksi total. Dapat dilihat bahwa akurasi model meningkat secara bertahap seiring dengan meningkatnya jumlah

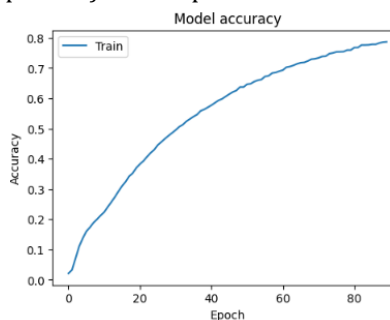
iterasi, menunjukkan bahwa model belajar dari data dan menjadi lebih baik dalam membuat prediksi yang benar.

Pada awal pelatihan, ada peningkatan akurasi yang signifikan, tetapi tingkat peningkatan melambat setelah sekitar 20 iterasi, menunjukkan bahwa model mendekati konvergensi. Akurasi model mencapai sekitar 80% setelah 90 iterasi, yang berarti bahwa model tersebut mampu memprediksi dengan benar sekitar 80% dari waktu pada data pelatihan. Meskipun hasil ini menunjukkan kinerja yang baik, penting untuk membandingkan akurasi pelatihan dengan validasi atau akurasi pengujian untuk memastikan bahwa overfitting tidak terjadi. Dengan demikian, sementara model menunjukkan tren peningkatan akurasi yang baik selama pelatihan, analisis lebih lanjut diperlukan untuk menilai kinerja model secara keseluruhan.

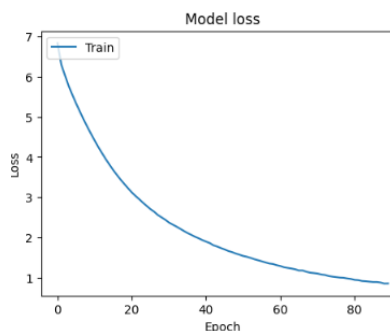
Tabel 2 Seleksi Model Terbaik

Metric	Value (Iterasi 30)	Value (Iterasi 60)	Value (Iterasi 90)
Loss	2.2836	1.2360	0.800
Accuracy	60.26%	71.13%	80.59%

Dari Gambar 5 dan Gambar 6, dapat dilihat bahwa grafik akurasi meningkat pesat pada 25 iterasi pertama, kemudian ketika mencapai pertengahan 25 dan 50, perlahan berhenti dan hanya naik turun dengan sedikit perubahan. Kemudian untuk grafik *loss function*, grafik menurun dengan cepat pada 50 iterasi pertama kemudian setelah itu menjadi lambat hingga mencapai *loss function* pada titik antara 1 hingga 0.



Gambar 5 Training Accuracy untuk 90 iterasi



Gambar 6 Training Loss untuk 90 iterasi

D. Evaluasi Kinerja

Untuk mengevaluasi efektivitas model yang dibangun, perlu dikembangkan fungsi yang dapat menghasilkan prediksi untuk kata berikutnya. Parameter mengacu pada input pengguna untuk kueri pencarian dan jumlah istilah yang diprediksi yang diinginkan. Berikut adalah ilustrasi format input dan output seperti yang ditunjukkan pada Gambar 7.

```
1 for i in range(4):
2     print(i+1, end=" ")
3     input_text = input("Masukkan teks awal: ")
4
5     sequence_length = max_sequence_len - 1
6     predicted_word = predict_next_word(model, tokenizer, input_text, sequence_length)
7     if predicted_word is None:
8         print("Tidak ada kata yang diprediksi.")
9     else:
10        print(i+1, end=" ")
11        print("Predicted next word:", predicted_word)
```

1. Masukkan teks awal: skizofrenia parah
1. Predicted next word: paranoid
2. Masukkan teks awal: suara bisikan
2. Predicted next word: masih
3. Masukkan teks awal: perasaan baik
3. Predicted next word: tenang
4. Masukkan teks awal: wahan curiga
4. Predicted next word: halusinasi

Gambar 7 Fungsi Next Word Prediction

Berdasarkan nilai perangkat dan parameter yang tercantum dalam Tabel 1 dan 2, kinerja keseluruhan dari model jaringan GRU yang diusulkan dievaluasi secara menyeluruh pada kumpulan data pengujian menggunakan 90 iterasi. Evaluasi ini melibatkan beberapa iterasi pelatihan dan pengujian untuk memastikan kekokohan model. Model ini disesuaikan untuk mengoptimalkan kemampuan prediksinya, dan sebagai hasilnya, ia mencapai akurasi prediksi yang terpuji sebesar 87,04%. Akurasi ini menunjukkan efektivitas model dalam menangkap pola dalam kumpulan data dan potensi penerapannya dalam skenario rekam medis elektronik.

IV. KESIMPULAN

Model GRU yang dikembangkan menunjukkan kinerja yang memuaskan pada data pelatihan dengan akurasi 80,59%. Hal ini menunjukkan bahwa model tersebut mampu mempelajari pola dan hubungan antar kata dalam rekam medis dengan baik. Penelitian menunjukkan bahwa GRU memiliki kemampuan untuk menangkap ketergantungan jangka panjang dalam data teks, yang sangat berguna dalam konteks memprediksi kata berikutnya (Cho et al., 2014). Selain itu, nilai loss function dalam data pelatihan cenderung menurun terus-menerus selama proses pelatihan, yang menunjukkan bahwa model tidak mengalami overfitting.

Pada data pengujian, model GRU menunjukkan akurasi 87,04%, sedikit lebih tinggi dari data pelatihan dalam kisaran yang dapat diterima. Hal ini menunjukkan bahwa model tersebut memiliki generalisasi yang baik terhadap data yang belum pernah terlihat sebelumnya.

Perbedaan kecil antara kinerja pada data pelatihan dan pengujian adalah indikator bahwa model tidak overfitting dan memiliki kemampuan generalisasi yang baik. Selain itu, nilai presisi dan penarikan masing-masing mencapai 88% dan 87%, menunjukkan bahwa model tersebut mampu membuat prediksi dengan tingkat kesalahan yang rendah.

UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih kepada Dekan Fakultas MIPA Universitas Tanjungpura, atas dukungan dan dorongan yang telah diberikan selama penelitian ini.

V. REFERENSI

- Arfianti, U. I., Novitasari, D. C. R., Widodo, N., Hafiyusholeh, Moh., & Utami, W. D. (2021). Sunspot Number Prediction Using Gated Recurrent Unit (GRU) Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(2), 141. <https://doi.org/10.22146/ijccs.63676>
- Boonstra, A., Versluis, A., & Vos, J. F. J. (2014). Implementing electronic health records in hospitals: A systematic literature review. *BMC Health Services Research*, 14(1), 370. <https://doi.org/10.1186/1472-6963-14-370>
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014, September 2). *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. arXiv. Retrieved from <http://arxiv.org/abs/1406.1078>
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014, December 11). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. arXiv. Retrieved from <http://arxiv.org/abs/1412.3555>
- Harrison, C. J., & Sidey-Gibbons, C. J. (2021). Machine learning in medicine: A practical introduction to natural language processing. *BMC Medical Research Methodology*, 21(1), 158. <https://doi.org/10.1186/s12874-021-01347-1>
- Krishnan, S., Magalingam, P., & Ibrahim, R. (2021). Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(6), 5467. <https://doi.org/10.11591/ijece.v11i6.pp5467-5476>
- Pierre, A. A., Akim, S. A., Semenyó, A. K., & Babiga, B. (2023). Peak Electrical Energy Consumption

- Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches. *Energies*, 16(12), 4739. <https://doi.org/10.3390/en16124739>
- Sidey-Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: A practical introduction. *BMC Medical Research Methodology*, 19(1), 64. <https://doi.org/10.1186/s12874-019-0681-4>
- Zhang, D., Yin, C., Zeng, J., Yuan, X., & Zhang, P. (2020). Combining structured and unstructured data for predictive models: A deep learning approach. *BMC Medical Informatics and Decision Making*, 20(1), 280. <https://doi.org/10.1186/s12911-020-01297-6>
- Zhang, Y., & Yang, Q. (2022). A Survey on Multi-Task Learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(12), 5586-5609. <https://doi.org/10.1109/TKDE.2021.3070203>