

**KLASIFIKASI *ELECTRONIC HEALTH RECORD* (EHR)
MENGUNAKAN *LONG SHORT TERM MEMORY* (LSTM) DAN *GATED
RECURRENT UNIT* (GRU)**

(Skripsi)

Oleh

ZIDA BUNGA SOBARA



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2023**

ABSTRACT

ELECTRONIC HEALTH RECORD (EHR) CLASSIFICATION USING LONG SHORT TERM MEMORY (LSTM) AND GATED RECURRENT UNIT (GRU)

By

ZIDA BUNGA SOBARA

The use of an Electronic Health Record (EHR) can generate a lot of information about a person's health digitally and can be used as the main consideration factor used by doctors to determine the next course of action for patients, whether they need to be hospitalized (in-patient care) or out-patient care. The importance of the EHR data requires the hospital to be responsive in determining further actions for patients. This study aims to classify EHR data using the LSTM and GRU methods. Both of these methods are the development of the Recurrent Neural Network (RNN) method. Based on the results of its accuracy, the LSTM method has a better accuracy value than the GRU method in classifying EHR data with an accuracy value of 75.57% and 75.34% respectively.

Keywords : Electronic Health Record, treatment classification, patient's laboratory test results, LSTM, and GRU.

ABSTRAK

KLASIFIKASI *ELECTRONIC HEALTH RECORD* (EHR) MENGUNAKAN *LONG SHORT TERM MEMORY* (LSTM) DAN *GATED RECURRENT UNIT* (GRU)

Oleh

ZIDA BUNGA SOBARA

Electronic Health Record (EHR) merupakan data medis yang dapat menghasilkan banyak informasi tentang kesehatan seseorang secara digital dan dapat dijadikan sebagai faktor pertimbangan utama yang digunakan oleh dokter dalam menentukan tindakan lanjutan bagi pasien, yaitu apakah pasien perlu dirawat di rumah sakit (rawat inap) atau rawat jalan. Pentingnya data EHR tersebut mengharuskan pihak rumah sakit untuk cepat tanggap dalam menentukan tindakan lanjutan bagi pasien. Penelitian ini bertujuan untuk melakukan klasifikasi data EHR dengan menggunakan metode LSTM dan GRU. Kedua metode tersebut merupakan perkembangan dari metode *Recurrent Neural Network* (RNN). Berdasarkan hasil akurasi, metode LSTM memiliki nilai akurasi yang lebih baik dibandingkan metode GRU dalam mengklasifikasikan data EHR dengan nilai akurasi masing-masing sebesar 75.57% dan 75.34%.

Kata kunci : Electronic Health Record, klasifikasi pelayanan, hasil tes laboratorium pasien, LSTM, dan GRU.

**KLASIFIKASI *ELECTRONIC HEALTH RECORD* (EHR)
MENGUNAKAN *LONG SHORT TERM MEMORY* (LSTM) DAN *GATED
RECURRENT UNIT* (GRU)**

Oleh

Zida Bunga Sobara

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi : **KLASIFIKASI *ELECTRONIC HEALTH RECORD* (EHR) MENGGUNAKAN *LONG SHORT TERM MEMORY* (LSTM) DAN *GATED RECURRENT UNIT* (GRU)**

Nama Mahasiswa : **Zida Bunga Sobara**

Nomor Pokok Mahasiswa : **1917031091**

Program Studi : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



1. Komisi Pembimbing

A handwritten signature in blue ink, appearing to be 'Dian Kurniasari', written over a white rectangular area.

Dian Kurniasari, S.Si., M.Sc.
NIP. 19690305 199603 2 001

A handwritten signature in blue ink, appearing to be 'Aang Nuryaman', written over a white rectangular area.

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

Ketua Jurusan Matematika

A handwritten signature in blue ink, appearing to be 'Aang Nuryaman', written over a white rectangular area.

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

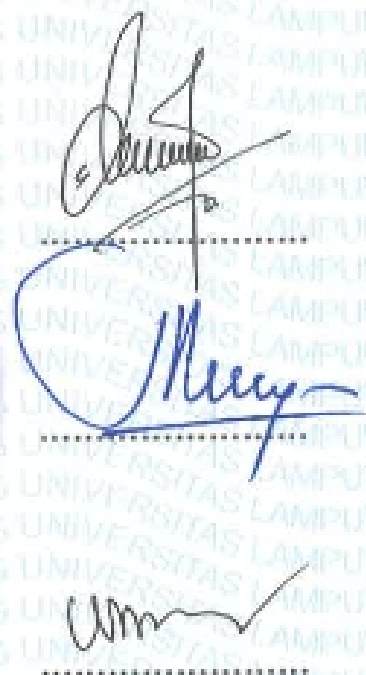
MENGESAHKAN

1. Tim Penguji

Ketua : Dian Kurniasari, S.Si., M.Sc.

Sekretaris : Dr. Aang Nuryaman, S.Si., M.Si.

Penguji Bukan Pembimbing : Ir. Warsono, M.S., Ph.D.



**2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002**

Tanggal Lulus Ujian Skripsi: 20 Juni 2023

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan dibawah ini:

Nama : **Zida Bunga Sobara**
Nomor Pokok Mahasiswa : **1917031091**
Jurusan : **Matematika**
Judul Skripsi : ***KLASIFIKASI *ELECTRONIC HEALTH RECORD* (EHR) MENGGUNAKAN *LONG SHORT TERM MEMORY* (LSTM) DAN *GATED RECURRENT UNIT* (GRU)***

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila di kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 20 Juni 2023

Penulis,



Zida Bunga Sobara
NPM. 1917031091

RIWAYAT HIDUP

Zida Bunga Sobara lahir di Kota Jakarta pada 30 November 2000. Penulis merupakan anak kedua dari dua bersaudara pasangan Bapak Drs. Misfatoni dan Ibu Iyam Mariyam. Penulis menempuh pendidikan pertamanya di Taman Kanak-Kanak Muhajirin pada tahun 2006-2007 dan melanjutkan pendidikan dasar di SDN 13 Pagi pada tahun 2007-2013. Selanjutnya, penulis melanjutkan jenjang pendidikannya di SMP Negeri 236 Jakarta pada tahun 2013-2016. Kemudian, penulis melanjutkan lagi jenjang pendidikannya di SMA Negeri 89 Jakarta pada tahun 2016-2019. Setelah itu penulis terdaftar sebagai mahasiswa Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) pada tahun 2019.

Selama menjadi mahasiswa, penulis aktif di beberapa kegiatan diantaranya:

1. Pada tahun 2019 penulis menjadi anggota magang bidang Kesekretariatan di UKMF Natural FMIPA Unila
2. Pada bulan Januari – Desember 2020 penulis menjadi anggota pengurus UKMF Natural FMIPA Unila.
3. Pada bulan Januari – Desember 2021 penulis menjadi Kepala Biro Kesekretariatan UKMF Natural FMIPA Unila.
4. Pada bulan Januari-Februari 2022 penulis melaksanakan Kerja Praktik di Badan Pendapatan Daerah Kabupaten Pandeglang.
5. Pada bulan Juni-Agustus 2022 penulis melaksanakan Kuliah Kerja Nyata di Pulau Meranti, Kecamatan Gunung Pelindung, Kabupaten Lampung Timur, Provinsi Lampung.

KATA INSPIRASI

“Maka milik Allahlah kehidupan akhirat dan kehidupan dunia”

(Q.S. An-Najm:25)

“Wahai orang-orang yang beriman!

Mohonlah pertolongan (kepada Allah) dengan sabar dan sholat. Sungguh, Allah beserta orang-orang yang sabar”

(Q.S. Al-Baqarah:153)

“Process on your steps must be appreciated even though it’s just say thanks to yourself”

(Penulis)

“The act of wanting to pursue something maybe even more precious than actually becoming that thing, so just being in the process itself is a prize”

(Mark Lee)

PERSEMBAHAN

Dengan mengucapkan Alhamdulillah dan rasa syukur kepada Allah SWT yang Maha Agung, saya persembahkan rasa terima kasih saya dengan hati yang tulus kepada:

Kedua Orang Tua Tersayang dan Kakak

Terima kasih kepada kedua orang tua yang telah tulus dan sabar memberikan doa, dukungan, motivasi, waktu, pengorbanan, dan kasih sayangnya kepada saya. Terimakasih atas segala ridho dan doa yang senantiasa kalian berikan demi mudah dan ringannya setiap perjalanan yang saya hadapi. Tak lupa juga saya ucapkan terimakasih banyak kepada kakak saya, Muhammad Syauqi yang selalu mendukung, mendoakan dan mempercayakan saya untuk tetap mampu berproses dan menggapai cita-cita.

Dosen Pembimbing dan Pembahas

Dosen-dosen yang dengan tulus memberikan ilmu, dukungan, pengalaman, motivasi, dan arahan yang sangat berharga kepada penulis.

Teman-teman yang telah kebersamai dan selalu mendukung serta mendoakan saya dalam setiap perjalanan hidup saya.

Almamater Tercinta Universitas Lampung

SANWACANA

Puji syukur saya haturkan kepada Allah SWT yang telah melimpahkan rahmat, nikmat dan karunia-Nya kepada penulis sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi *Electronic Health Record* (EHR) Menggunakan *Long Short Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU)”. Dalam kesempatan ini, penulis ingin mengucapkan terima kasih kepada:

1. Ibu Dian Kurniasari, S.Si., M.Sc., selaku Dosen Pembimbing I yang telah ikhlas memberikan bimbingan, bantuan, motivasi, dan sarannya kepada penulis dalam proses penyusunan skripsi ini.
2. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Dosen Pembimbing II dan Ketua Jurusan Matematika FMIPA Universitas Lampung yang telah memberikan arahan, kritik, dan sarannya kepada penulis dalam proses penelitian dan penyusunan skripsi ini.
3. Bapak Ir. Warsono, M.S., Ph.D., selaku Dosen Pembahas yang telah memberikan masukan, kritik, dan sarannya kepada penulis dalam proses penelitian dan penyusunan skripsi ini.
4. Bapak Amanto, S.Si., M.Si., (Alm) selaku Dosen yang telah memberikan arahan, kritik, dan sarannya kepada penulis dalam proses penelitian dan penyusunan skripsi ini.
5. Ibu Dr. Fitriani, S.Si., M.Sc., selaku Dosen Pembimbing Akademik yang telah memberikan bimbingan dan saran kepada penulis selama mengemban bangku perkuliahan.
6. Kedua orang tuaku, Bapak Drs. Misfatoni dan Ibu Iyam Mariyam, serta kakak yang senantiasa memberikan dukungan, semangat, motivasi, dan doa yang tiada henti demi kemudahan dan kelancaran penulis dalam penulisan

skripsi ini.

7. Seluruh dosen, staff, dan karyawan Jurusan Matematika FMIPA Universitas Lampung yang telah memberikan ilmu dan pengalaman yang berharga bagi penulis.
8. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
9. Anggun, Astina, berlian, mba linda, safitri, mega, siti, tika, nida, dan mba widya, yang selalu memberikan dukungan dan bantuan kepada penulis.
10. Kak Tiway yang sudah menemani, memotivasi, dan menghibur penulis.
11. Jaehyun, Mark, ilichil, Dipa, Lando, dan anggota NCT lainnya yang sudah memberikan semangat, motivasi, dan hiburan kepada penulis.
12. Teman-teman seperjuangan skripsi (Azza, Anin, Fiqih, Lisna, Silvi, Tiara, Dea, Mia, dan Puter) yang selalu memberikan dukungan dan bantuan kepada penulis.
13. Teman-teman Matematika 2019 dan UKMF Natural FMIPA Unila yang telah menjadi rekan seperjuangan selama menempuh bangku perkuliahan.
14. Semua pihak terkait yang telah memberikan pengalaman dan dukungan yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini jauh dari kata sempurna dan masih terdapat banyak kekurangan baik dalam penyajian maupun teknik penulisan. Oleh sebab itu, saran dan kritikan yang membangun senantiasa penulis harapkan demi menyempurnakan skripsi ini.

Bandar Lampung, 20 Juni 2023
Penulis,

Zida Bunga Sobara

DAFTAR ISI

	Halaman
DAFTAR TABEL	iii
DAFTAR GAMBAR	iv
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian.....	4
1.3 Manfaat.....	4
II. TINJAUAN PUSTAKA	5
2.1 Klasifikasi.....	5
2.2 <i>Electronic Health Record</i> (EHR)	5
2.3 <i>Preprocessing Data</i>	6
2.4 <i>Deep Learning</i>	7
2.4.1 <i>Recurrent Neural Network</i> (RNN).....	8
2.4.2 Fungsi Aktivasi.....	9
2.4.3 <i>Long Short Term Memory</i> (LSTM)	11
2.4.4 <i>Gated Recurrent Unit</i> (GRU)	15
2.5 <i>Hyperparameter</i>	17
2.6 Evaluasi Model.....	18
2.6.1 <i>Confusion Matrix</i>	19
III. METODOLOGI PENELITIAN	21
3.1 Waktu dan Tempat Penelitian	21
3.2 Data Penelitian.....	21
3.3 Metode Penelitian	23
IV. HASIL DAN PEMBAHASAN	25
4.1 <i>Input Data</i>	25
4.2 Seleksi Variabel.....	26
4.3 Visualisasi Data	27
4.4 <i>Preprocessing Data</i>	29

4.5	<i>Splitting Data</i>	30
4.6	Klasifikasi Menggunakan Model <i>Long Short Term Memory</i> (LSTM). 30	
4.6.1	Membangun Model LSTM	30
4.6.2	<i>Hypertuning Hyperparameter LSTM</i>	31
4.6.3	Prediksi Menggunakan Model LSTM	32
4.7	Evaluasi Model LSTM	33
4.8	Klasifikasi Menggunakan Model <i>Gated Recurrent Unit</i> (GRU).....	36
4.8.1	Membangun Model GRU	36
4.8.2	<i>Hypertuning Hyperparameter GRU</i>	36
4.8.3	Prediksi Menggunakan Model GRU	38
4.9	Evaluasi Model GRU	39
4.10	Perbandingan Hasil Klasifikasi Antara Model LSTM dan GRU.....	41
V.	KESIMPULAN	43
	DAFTAR PUSTAKA	44
	LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. <i>Confusion Matrix</i>	19
2. <i>Data Electronic Health Record (EHR)</i>	22
3. Jumlah Pasien Data EHR	28
4. Keberadaan <i>Missing Value</i>	29
5. Hasil <i>Hypertuning</i> Model LSTM.....	32
6. <i>Classification Report</i> Model LSTM	35
7. Hasil <i>Hypertuning</i> Model GRU	37
8. <i>Classification Report</i> Model GRU.....	40
9. Perbandingan Hasil Klasifikasi Model LSTM dan GRU dengan Data Testing	41

DAFTAR GAMBAR

Gambar	Halaman
1. Arsitektur RNN.	8
2. Grafik Fungsi Sigmoid.	9
3. Grafik Fungsi Tanh.	11
4. Arsitektur LSTM.	12
5. Arsitektur GRU.	15
6. Diagram Alir Metode LSTM dan GRU.	24
7. Data EHR.	25
8. Data EHR Setelah Pelabelan.	26
9. <i>Syntax</i> Pendefinisian Variabel pada Data EHR.	26
10. <i>Feature Variable</i>	27
11. <i>Predicted Variable</i>	27
12. <i>Bar Chart Source</i> Data EHR.	28
13. Pengecekan Data yang Berduplikat.	29
14. <i>Hypertuning</i> Model LSTM.	31
15. Prediksi Model LSTM.	32
16. Grafik Akurasi Model LSTM Terbaik.	33
17. <i>Confusion Matrix</i> Model LSTM.	34
18. <i>Hypertuning</i> Model GRU.	37
19. Prediksi Model GRU.	38

20. Grafik Akurasi Model GRU Terbaik.	38
21. <i>Confusion Matrix</i> Model GRU.....	39

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Electronic Health Record (EHR) merupakan data medis yang bersumber dari pasien. Semua data bisa dikumpulkan untuk menghasilkan informasi yang mencakup rincian riwayat medis seperti tes diagnostik, obat-obatan dan rencana perawatan, rekaman imunisasi, alergi, gambar radiologi, dan perangkat sensor multivariat (Tutuko, dkk., 2020). Semua informasi dari EHR akan memberikan pemahaman yang berharga tentang sistem manajemen penyakit. Selain itu, penggunaan EHR dapat menghasilkan banyak informasi tentang kesehatan seseorang secara digital. Dataset ini adalah faktor pertimbangan utama yang digunakan oleh dokter untuk menentukan tindakan lanjutan bagi pasien, yaitu apakah pasien perlu dirawat di rumah sakit (rawat inap) atau rawat jalan.

Di era pandemi, tentunya peran pelayanan kesehatan sangatlah penting. Namun saat pandemi mulai mereda pun peran pelayanan kesehatan juga masih menjadi suatu hal yang sangat penting bagi masyarakat. Pelayanan kesehatan adalah salah satu industri yang paling kompleks karena mencakup banyak pemangku kepentingan, berbagai alat, dan juga teknologi (Mavrogiorgou, dkk., 2019). Sejalan dengan meningkatnya pertumbuhan populasi di Indonesia, jumlah keberadaan rumah sakit juga meningkat. Meningkatnya jumlah rumah sakit di Indonesia tentunya harus sejalan dengan kualitas pelayanan kesehatan yang diberikan. Kualitas pelayanan kesehatan yang dimaksud adalah seperti peningkatan fasilitas atau peningkatan efisiensi pelayanan terhadap pasien.

Seiring dengan pentingnya kualitas pelayanan rumah sakit bagi masyarakat, hal ini mengharuskan rumah sakit untuk tanggap dalam melayani pasien, salah satunya adalah menentukan apakah pasien harus menjalani rawat inap atau rawat jalan. Perbedaan diantara rawat inap dan rawat jalan adalah pada seberapa lama pasien harus tinggal di fasilitas tempat mereka menjalani prosedur medis.

Sudah ada beberapa metode yang digunakan untuk mengklasifikasikan data EHR, diantaranya dalam penelitian Sadikin & Nurhaida (2021), digunakan beberapa metode seperti metode Decision Tree yang menghasilkan akurasi 63.39%, Gaussian Naïve Bayes menghasilkan akurasi 68.95%, dan Random Forest yang memberikan nilai akurasi sebesar 71.58%. Kemudian penelitian Rosita, dkk. (2022), melakukan prediksi apakah pasien perlu dirawat inap atau rawat jalan dengan menggunakan beberapa teknik *mechine learning* dan dihasilkan bahwa penggunaan *neural network* menghasilkan *f1-score* paling tinggi yaitu sebesar 67%.

Qiao, dkk. (2018), melakukan prediksi kunjungan *Emergency Room* (ER) berdasarkan data EHR menggunakan metode Linear Regression, XGBoost, dan RNN kemudian didapatkan bahwa XGBoost memiliki hasil yang lebih baik dengan nilai AUC sebesar 0.6974. Haris, dkk. (2021), melakukan klasifikasi suara detak jantung dengan LSTM 128 *neuron* dan GRU 128 *neuron* dan dihasilkan bahwa GRU memiliki presisi yang lebih baik sebesar 97%. Kemudian Handari, dkk. (2021), membandingkan *Elman Neural Network* (ENN), LSTM, dan GRU dalam memprediksi kasus demam berdarah di DKI Jakarta dan dapat disimpulkan bahwa LSTM memiliki hasil prediksi yang paling baik dengan nilai RMSE sebesar 0.079. Pardede & Raspati (2021), melakukan klasifikasi menggunakan GRU dalam mendeteksi *Obstructive Sleep Apnea* (OSA) menggunakan sinyal electrocardiogram (ECG) serta mengukur kinerja sistemnya berdasarkan *confusion matrix* dan diperoleh model dengan nilai akurasi sebesar 83.92%.

Salah satu perkembangan metode dalam melakukan klasifikasi adalah metode *deep learning*, metode ini merupakan bagian dari *machine learning* yang berkaitan dengan algoritma dimana cara kerja dari algoritma ini meniru struktur dan fungsi otak yang disebut jaringan saraf tiruan. Salah satu metode dari *deep learning* adalah *Recurrent Neural Network* (RNN). Hasil modifikasi dari metode RNN yaitu *Long Short Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU).

Recurrent Neural Network (RNN) memperbolehkan *output* sebelumnya menjadi *input* saat ini dalam *states*. RNN memiliki kelemahan yaitu tidak mampu dalam menyimpan terlalu banyak informasi pada proses *learning* nya. Masalah ini dapat diselesaikan dengan teknik LSTM dan GRU. Menurut Putra & Hendry (2022), LSTM adalah sebuah algoritma pengembangan yang ada untuk mengatasi permasalahan pada RNN yaitu *vanishing gradient*. Hal ini sering terjadi pada data yang memiliki *sequence* yang panjang sehingga RNN tidak dapat menangkap *long term dependencies*.

Sementara GRU adalah modifikasi RNN yang sedikit lebih sederhana daripada LSTM. Secara signifikan, sifat GRU yang sedikit lebih sederhana daripada LSTM membuat GRU lebih cepat dalam menghitung beberapa set data. Namun tidak dapat dipastikan mana yang lebih baik diantara kedua metode tersebut. Sama seperti LSTM, GRU hadir untuk menyelesaikan masalah pada RNN mengenai *vanishing gradient* (Shen, dkk., 2018). Selain itu, metode LSTM dan GRU mempunyai kemampuan untuk menganalisis berbagai jenis data seperti data populasi, data laboratorium, gambar medis atau catatan dokter.

Oleh karena itu, penelitian ini memanfaatkan data EHR pasien dari hasil pemeriksaan laboratorium yang diambil dari salah satu rumah sakit swasta di Indonesia untuk diklasifikasi menggunakan metode LSTM dan GRU sehingga diperoleh model terbaik dari masing-masing metode berdasarkan hasil akurasi.

1.2 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

1. Membuat model dengan metode LSTM dan GRU untuk melakukan klasifikasi jenis pelayanan pasien rumah sakit.
2. Membandingkan *running time* yang diperlukan model LSTM dan GRU dalam melakukan klasifikasi data *Electronic Health Record* (EHR).
3. Mengevaluasi model terbaik dari masing-masing metode yang telah dirancang.
4. Membandingkan hasil akurasi diantara kedua metode dalam penelitian klasifikasi jenis pelayanan pasien rumah sakit.

1.3 Manfaat

1. Memperoleh model dan nilai akurasi dari masing-masing metode LSTM dan GRU dalam melakukan klasifikasi data *Electronic Health Record* (EHR) pasien dari hasil pemeriksaan laboratorium yang diambil dari salah satu rumah sakit swasta di Indonesia.
2. Mengetahui perbedaan *running time* yang diperlukan model LSTM dan GRU dalam melakukan klasifikasi data *Electronic Health Record* (EHR).
3. Penelitian ini dapat dijadikan bahan referensi bagi para tenaga medis dalam menentukan tindakan lanjutan bagi pasien, yaitu apakah perlu dirawat inap atau rawat jalan.

II. TINJAUAN PUSTAKA

2.1 Klasifikasi

Klasifikasi adalah pengelompokan yang sistematis dari sejumlah obyek, gagasan atau benda-benda lain ke dalam kelas atau golongan tertentu berdasarkan ciri-ciri yang sama (Hamakonda & Tairas, 2008). Menurut Han, dkk. (2012), konsep klasifikasi merupakan bagian dari teknik data *mining* yang memiliki pekerjaan utama melakukan analisis prediksi. Pada klasifikasi, data *training* memiliki label kelas yang telah dibuat oleh pakar. Kemudian berdasarkan pembelajaran pada data *training* yang telah berlabel tersebut, model klasifikasi dapat memprediksi label untuk data baru.

2.2 *Electronic Health Record (EHR)*

Menurut Tutuko, dkk. (2020), *Electronic Health Record (EHR)* merupakan data medis yang bersumber dari pasien dan bisa dikumpulkan untuk menghasilkan informasi yang mencakup rincian riwayat medis seperti tes diagnostik, obat-obatan dan rencana perawatan, rekaman imunisasi, alergi, gambar radiologi, dan perangkat sensor multivariat. Informasi kesehatan yang berkaitan dengan pasien dikumpulkan secara sistematis dan disimpan secara elektronik dalam format digital.

Salah satu fitur utama EHR adalah informasi kesehatannya dapat dibuat dan dikelola oleh penyedia resmi dalam format digital yang dapat dibagikan dengan

penyedia lain di lebih dari satu organisasi layanan kesehatan. EHR dibangun untuk berbagi informasi dengan penyedia dan organisasi layanan kesehatan lainnya seperti laboratorium, spesialis, fasilitas pencitraan medis, apotek, fasilitas darurat, dan klinik sekolah dan tempat kerja. Sehingga EHR berisi informasi dari semua ahli klinis atau dokter yang terlibat dalam perawatan pasien.

EHR merupakan faktor pertimbangan utama yang digunakan oleh dokter untuk menentukan tindakan lanjutan bagi pasien, yaitu apakah pasien perlu dirawat di rumah sakit (rawat inap) atau rawat jalan. Penting bagi pasien untuk memahami perbedaan diantara kedua tipe pelayanan tersebut karena berkaitan dengan lamanya pasien untuk tetap dalam fasilitas kesehatan dan juga biaya prosedur medis yang harus dibayarkan.

Rawat inap mensyaratkan pasiennya untuk tetap dalam fasilitas kesehatan rumah sakit setidaknya untuk satu malam. Selama waktu ini, mereka tetap berada di bawah pengawasan perawat atau dokter. Sedangkan pasien yang menerima rawat jalan tidak perlu menghabiskan malamnya di rumah sakit. Mereka bebas untuk meninggalkan rumah sakit setelah prosedur kesehatan selesai. Dalam beberapa kasus, pasien butuh untuk menunggu sampai penggunaan anestesi hilang atau untuk memastikan bahwa tidak ada komplikasi yang muncul. Selama tidak ada komplikasi serius, pasien tidak harus menghabiskan malamnya untuk diawasi oleh tenaga medis.

2.3 *Preprocessing Data*

Salah satu langkah yang perlu dilakukan dalam *preprocessing* data adalah pengecekan keberadaan *missing value*. Langkah lanjutan ketika ditemukan adanya *missing value* adalah melakukan *data cleaning*. Menurut Handari, dkk. (2021), *preprocessing* data terdiri dari imputasi *mean*, perubahan data target, dan normalisasi data. Imputasi *mean* hanya diperlukan ketika terdapat data kosong atau data *outlier*. *Features variable* biasanya dinormalisasi sebelum diproses

lebih lanjut di jaringan saraf. Proses normalisasi data bermanfaat untuk proses pelatihan.

2.4 *Deep Learning*

Deep learning adalah bagian dari pembelajaran mesin (*machine learning*) yang berkaitan dengan algoritma dimana cara kerja dari algoritma ini meniru struktur dan fungsi otak yang disebut jaringan saraf tiruan. Menurut Francois, dkk. (2018), kata “*deep*” di dalam istilah *deep learning* berkaitan dengan jumlah layer dari model *deep learning* yang dipergunakan.

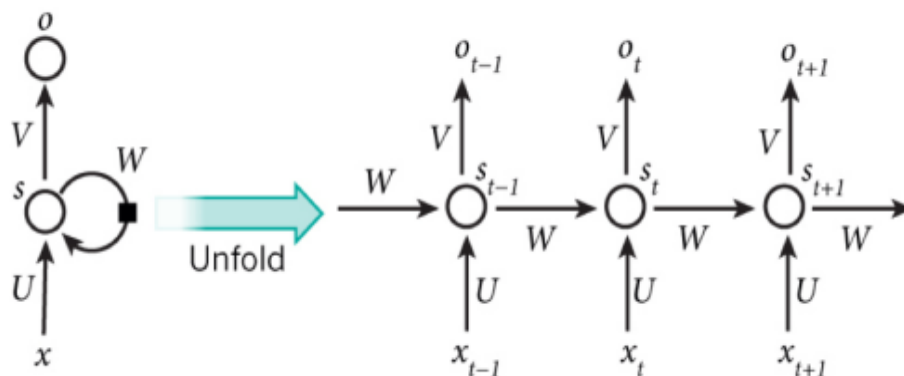
Menurut Heryadi & Irwansyah (2020), beberapa sifat dari *deep learning*, yaitu:

- 1) *Simplicity*: data yang digunakan sebagai *input* proses pembelajaran tidak membutuhkan proses rekayasa fitur sebelumnya, misalnya ekstraksi fitur. Hal ini menjadikan proses pembelajaran *deep learning* menjadi lebih sederhana.
- 2) *Scability*: algoritma pembelajaran yang bersifat iteratif menggunakan sejumlah *batch* data memungkinkan pembelajaran model menggunakan *dataset* berbagai *volume* bahkan berskala besar.
- 3) *Versatility (adaptability)*: proses pembelajaran model *deep learning* tidak selalu harus dilakukan dari awal. Sebuah model *deep learning* hasil pembelajaran dapat di *training* kembali menggunakan data *training* yang baru. Hal ini menjadikan model *deep learning* sebagai sebuah model yang mampu untuk terus menerus melakukan pembelajaran secara *online*.
- 4) *Reusability*: sebuah model hasil yang telah di *training* menggunakan *dataset* berskala besar dapat digunakan untuk melakukan tugas menggunakan *dataset* yang berskala lebih kecil. Kemampuan ini disebut *transfer learning*, menghemat waktu pembelajaran ulang.

2.4.1 Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) atau disebut juga jaringan umpan balik merupakan metode yang didalamnya terdapat *loop* yang mengakomodasi *output* pada *network* untuk menjadi *input* pada *network* tersebut yang kemudian akan digunakan untuk menghasilkan *output* baru (Putra & Hendry, 2022). Memori internal dipertahankan oleh RNN untuk mengenali pola atau *pattern* yang dinamis. RNN memiliki arsitektur yang mana *hidden layer* nya memperbolehkan *output* sebelumnya menjadi *input* saat ini dalam *states*.

Proses para RNN akan dipanggil berulang-ulang untuk mengerjakan masukan yaitu data sekuensial. Mengklasifikasikan data *time series* dan sekuensial merupakan ciri khas dari RNN. Data *time series* adalah data yang digabungkan menurut urutan waktu dalam rentang tertentu, sedangkan data sekuensial merupakan suatu sampel data yang diproses secara terurut dan setiap urutan berhubungan satu sama lain



Gambar 1. Arsitektur RNN.
(Sumber Singh, dkk., 2018)

Simbol x_t pada gambar 1 adalah input pada tiap proses atau *time step*, simbol S_t merupakan *hidden state* atau *memory* pada tiap *time step* t yang berfungsi menyimpan hasil perhitungan yang telah dilakukan. *Hidden state* dihitung

berdasarkan *hidden state* sebelumnya dan berdasarkan input pada *current state* (keadaan saat ini) sedangkan O_t merupakan *output* pada *time step* t .

2.4.2 Fungsi Aktivasi

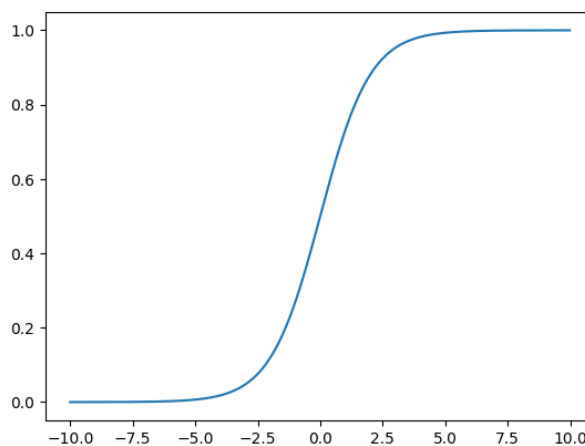
Fungsi aktivasi digunakan untuk menghitung nilai *output* berdasarkan *input* dan bobot pada neuron. Berikut merupakan fungsi aktivasi yang digunakan pada metode LSTM dan GRU:

1. Sigmoid

Sigmoid akan menerima angka tunggal dan mengubah nilai x menjadi sebuah nilai yang memiliki *range* antara 0 hingga 1 (Sitepu & Sigiro, 2021). Fungsi aktivasi sigmoid dirumuskan sebagai berikut:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

Adapun grafik yang dihasilkan dari fungsi sigmoid yaitu:



Gambar 2. Grafik Fungsi Sigmoid.

Turunan fungsi sigmoid diperoleh sebagai berikut:

$$\begin{aligned}
 f'(x) &= \frac{u(x)}{v(x)} = \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2} \\
 \frac{d}{dx} \sigma(x) &= \frac{d}{dx} \frac{1}{1 + e^{-x}} \\
 &= \frac{e^{-x}}{(1 + e^{-x})^2} \\
 &= \frac{1 - 1 + e^{-x}}{(1 + e^{-x})^2} \\
 &= \frac{1 + e^{-x}}{(1 + e^{-x})^2} - \frac{1}{(1 + e^{-x})^2} \\
 &= \frac{1}{1 + e^{-x}} - \frac{1}{(1 + e^{-x})^2} \\
 &= \frac{1}{1 + e^{-x}} \left(1 - \frac{1}{1 + e^{-x}} \right) \\
 &= \sigma(x)(1 - \sigma(x)) \tag{2.2}
 \end{aligned}$$

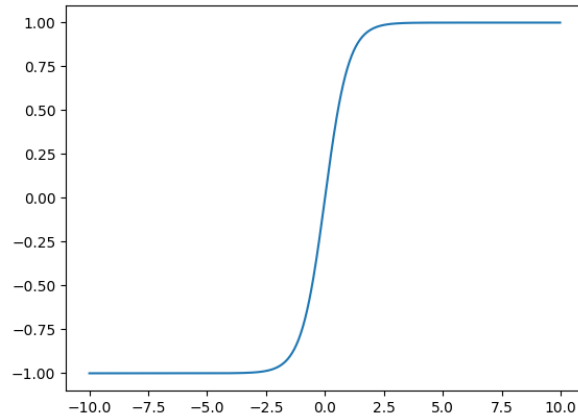
2. Tanh

Fungsi tanh atau disebut juga dengan tangen hiperbolik mempunyai *output* dengan rentang nilai dari -1 sampai 1 . Fungsi tanh memiliki grafik yang dapat dikatakan hampir sama seperti grafik fungsi sigmoid. Akan tetapi, fungsi tanh memiliki rentang nilai yang lebih luas dibandingkan dengan fungsi sigmoid.

Fungsi tanh dirumuskan sebagai berikut:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.3}$$

Adapun grafik yang dihasilkan dari fungsi tanh yaitu:



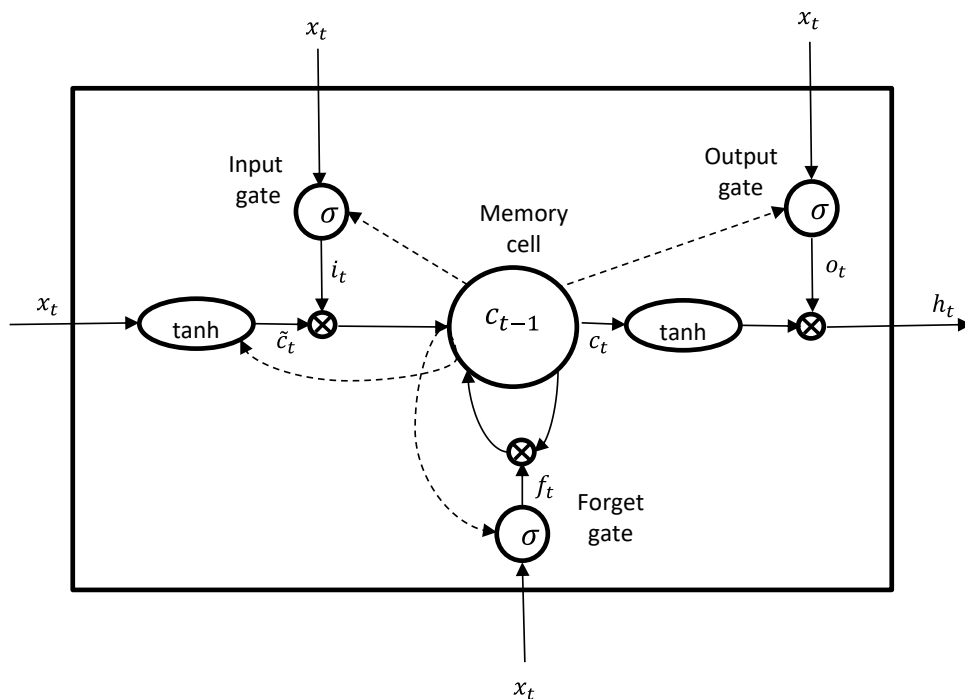
Gambar 3. Grafik Fungsi Tanh.

Turunan fungsi tanh diperoleh sebagai berikut:

$$\begin{aligned} f'(x) &= \frac{u(x)}{v(x)} = \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2} \\ \frac{d}{dx} \tanh(x) &= \frac{d}{dx} \frac{\sinh(x)}{\cosh(x)} \\ &= \frac{\cosh^2(x) - \sinh^2(x)}{\cosh^2(x)} \\ &= \frac{\cosh^2(x)}{\cosh^2(x)} - \frac{\sinh^2(x)}{\cosh^2(x)} \\ &= 1 - \tanh^2(x) \end{aligned} \tag{2.4}$$

2.4.3 Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) merupakan RNN yang telah dimodifikasi dengan menambahkan *memory cell* yang dapat menyimpan informasi untuk jangka waktu yang lama (Manaswi, 2018).



Gambar 4. Arsitektur LSTM.
(Sumber Sholeh & Hidayat, 2022)

Model LSTM mempunyai pemrosesan yang berbeda dengan model RNN biasa. Pada LSTM terdapat tambahan sinyal yang diberikan dari satu langkah waktu ke langkah waktu berikutnya, yaitu *memory cell*, direpresentasikan dengan simbol C_t . Selain itu, perbedaan antara LSTM dan RNN adalah ditemukannya penggunaan dari kombinasi fungsi aktivasi yang menentukan apakah informasi harus disimpan atau dihapus.

Metode LSTM ini mempunyai tiga fungsi *gate* yang dinotasikan oleh *input gate* i_t , *forget gate* f_t , dan *output gate* o_t . *Forget gate* adalah langkah awal yang berfungsi untuk memutuskan informasi apa yang akan disimpan dan dibuang dari *cell state* sebelumnya (c_{t-1}). *Forget gate* dihitung dari gabungan *input* sekarang dan *hidden state* sebelumnya. Dengan fungsi sigmoid, *gate* ini akan menghasilkan nilai antara 0 dan 1. *Forget gate* yang bernilai 0 menunjukkan bahwa *cell state* pada *time step* sebelumnya dilupakan, sedangkan *forget gate* yang bernilai 1 menunjukkan bahwa *cell state* pada *time step* sebelumnya lanjut ke *time step* sekarang. Berikut persamaan untuk *forget gate*:

$$f_t = \sigma (W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \quad (2.5)$$

dimana:

- f_t : *forget gate* pada waktu t
- σ : fungsi aktivasi sigmoid
- W_f : bobot *forget gate* yang terkoneksi dengan *input*
- x_t : nilai *input* pada waktu t
- U_f : bobot *forget gate* yang terkoneksi pada *output* sebelumnya
- h_{t-1} : *hidden state* pada waktu t-1
- b_f : nilai bias pada *forget gate*

Selanjutnya adalah *input gate* yang melambangkan seberapa banyak *input* yang akan ditambahkan ke *cell state* baru. *Input gate* terdiri dari dua bagian. Bagian pertama akan memilah dan menentukan informasi tertentu yang akan diperbarui ke bagian *cell state* dengan menggunakan fungsi aktivasi sigmoid dengan persamaan sebagai berikut:

$$i_t = \sigma (W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \quad (2.6)$$

dimana:

- i_t : *input gate* pada waktu t
- σ : fungsi aktivasi sigmoid
- W_i : bobot *input gate* yang terkoneksi dengan *input*
- x_t : nilai *input* pada waktu t
- U_i : bobot *input gate* yang terkoneksi pada *output* sebelumnya
- h_{t-1} : *hidden state* pada waktu t-1
- b_i : nilai bias pada *input gate*

Pada bagian kedua akan dibentuk kandidat vektor baru menggunakan fungsi aktivasi tanh yang dapat mengontrol seberapa banyak informasi baru yang ditambahkan pada bagian *cell state* dengan persamaan sebagai berikut:

$$\tilde{c}_t = \tanh (W_{\tilde{c}} \cdot x_t + U_{\tilde{c}} \cdot h_{t-1} + b_{\tilde{c}}) \quad (2.7)$$

dimana:

- \tilde{c}_t : nilai baru yang ditambahkan ke *cell state* pada waktu t.
- $W_{\tilde{c}}$: bobot *cell state* yang terkoneksi dengan *input*

- x_t : nilai *input* pada waktu t
- $U_{\tilde{c}}$: bobot *cell state* yang terkoneksi pada *output* sebelumnya
- h_{t-1} : *hidden state* pada waktu t-1
- $b_{\tilde{c}}$: nilai bias pada *cell state*

Setelah itu, proses memperbarui informasi lama c_{t-1} ke informasi baru c_t dengan persamaan sebagai berikut:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (2.8)$$

dimana:

- c_t : *cell state* pada waktu t
- f_t : *forget gate* pada waktu t
- c_{t-1} : *cell state* pada waktu t-1
- i_t : *input gate* pada waktu t
- \tilde{c}_t : nilai baru yang ditambahkan ke *cell state* pada waktu t.

Output gate merupakan proses yang memutuskan informasi apa yang akan dihasilkan dengan menjalankan sigmoid untuk menghasilkan nilai *output* pada *hidden state* dan menempatkan *cell state* pada tanh. Setelah menghasilkan nilai *output* sigmoid dan nilai *output* tanh, kedua hasil aktivasi tersebut dilakukan perkalian untuk menghasilkan nilai *output* akhir dari keseluruhan perhitungan pada proses LSTM.

Berikut persamaan yang digunakan pada *output gate*:

$$o_t = \sigma (W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \quad (2.9)$$

$$h_t = o_t \odot \tanh(c_t) \quad (2.10)$$

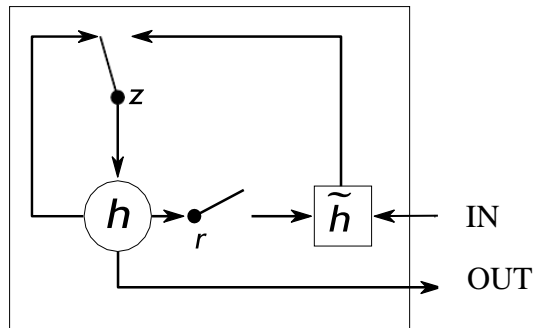
dimana:

- o_t : *output gate* pada waktu t
- σ : fungsi aktivasi sigmoid
- W_o : bobot *output gate* yang terkoneksi dengan *input*
- x_t : nilai *input* pada waktu t
- U_o : bobot *output gate* yang terkoneksi pada *output* sebelumnya
- h_{t-1} : *hidden state* pada waktu t-1
- b_o : nilai bias pada *output gate*

- h_t : *hidden state* pada waktu t
- c_t : *cell state* pada waktu t

2.4.4 Gated Recurrent Unit (GRU)

Variasi dari RNN yang lebih sederhana daripada LSTM adalah GRU. Terdapat kesamaan peran dalam jaringan antara GRU dengan LSTM. Perbedaan mendasar antara GRU dan LSTM adalah GRU tidak mempunyai *cell state* (Wicaksono, dkk., 2022). Sehingga GRU hanya memiliki *hidden state* yang lanjut ke *timestep* berikutnya. GRU mempunyai dua *gate* yaitu *reset gate* dan *update gate*.



Gambar 5. Arsitektur GRU.
(Sumber Hastomo, dkk., 2022)

Reset gate menentukan seberapa banyak informasi dari masa lampau yang harus diabaikan. Nilai dari *reset gate* dihasilkan dengan persamaan sebagai berikut:

$$r_t = \sigma (W_r \cdot x_t + U_r \cdot h_{t-1} + b_r) \quad (2.11)$$

dimana:

- r_t : *reset gate*
- σ : fungsi sigmoid
- W_r : bobot *reset gate* yang terkoneksi dengan *input*
- x_t : nilai *input* pada waktu t
- U_r : bobot *reset gate* yang terkoneksi pada *output* sebelumnya
- h_{t-1} : *hidden state* pada waktu $t-1$
- b_r : nilai bias pada *reset gate*

Sementara itu, *update gate* berfungsi mengontrol aliran informasi dari aktivasi sebelumnya dan menambahkan informasi baru juga. *Update gate* bertujuan untuk menentukan seberapa banyak informasi dari *time step* sebelumnya yang harus dibawa ke *timestep* berikutnya dengan persamaan sebagai berikut:

$$z_t = \sigma (W_z \cdot x_t + U_z \cdot h_{t-1} + b_z) \quad (2.12)$$

dimana:

z_t : *update gate*

σ : fungsi sigmoid

W_z : bobot *update gate* yang terkoneksi dengan *input*

x_t : nilai *input* pada waktu t

U_z : bobot *update gate* yang terkoneksi pada *output* sebelumnya

h_{t-1} : *hidden state* pada waktu t-1

b_z : nilai bias pada *update gate*

Candidates hidden state \tilde{h}_t menggunakan hasil dari *reset gate* untuk menyimpan informasi yang relevan dari masa lampau. Nilai dari \tilde{h}_t dihasilkan dengan persamaan berikut:

$$\tilde{h}_t = \tanh(W_h \cdot x_t + U_h(h_{t-1} \odot r_t) + b_h) \quad (2.13)$$

dimana:

\tilde{h}_t : *candidates hidden state*

W_h : bobot *candidates hidden state* yang terkoneksi pada *input*

x_t : nilai *input* pada waktu t

U_h : bobot *candidates hidden state* yang terkoneksi pada *output* sebelumnya

h_{t-1} : variabel *hidden state* pada waktu t-1

r_t : *reset gate*

b_h : nilai bias pada *candidates hidden state*

Ketika r_t mendekati satu, \tilde{h}_t menyimpan sementara nilai *output*, ketika r_t mendekati nol, \tilde{h}_t cenderung hanya menerima proses *input*. Sehingga, \tilde{h}_t menyimpan *output* pada *timesetp* t untuk lanjut ke tahap GRU berikutnya.

Selanjutnya akan dihasilkan nilai *output* akhir dari keseluruhan perhitungan pada proses GRU yaitu dengan persamaan sebagai berikut:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2.14)$$

dimana:

h_t : *output* atau *hidden state*

z_t : *update gate*

h_{t-1} : variabel *hidden state* pada waktu t-1

b_z : nilai bias pada *update gate*

\tilde{h}_t : *candidates hidden state*

2.5 *Hyperparameter*

Hyperparameter adalah variabel yang menentukan bagaimana sebuah model dilatih. Dalam kasus *neural networks*, beberapa *hyperparameter* yang umum yaitu *batch size*, jumlah *epoch*, *learning rate*, jumlah *hidden layers*, jumlah *neurons* di setiap layer, dan fungsi aktivasi. *Batch size* merupakan jumlah sampel data yang disebarkan ke *neural network*. *Epoch* adalah *hyperparameter* yang menentukan berapa kali algoritma *deep learning* bekerja melewati seluruh dataset. Nilai *learning rate* yang semakin besar akan mengakibatkan *training* berjalan semakin cepat. Selain itu, nilai *learning rate* juga mempengaruhi ketelitian jaringan. Nilai *learning rate* yang semakin besar akan mengakibatkan ketelitian jaringan semakin berkurang. Sebaliknya, nilai *learning rate* yang semakin kecil akan mengakibatkan ketelitian jaringan semakin bertambah dengan konsekuensi proses *training* akan memakan waktu yang semakin lama.

Sementara *hypertuning* adalah proses pembentukan model dengan menguji setiap kombinasi *hyperparameter* untuk mendapatkan *hyperparameter* terbaik sehingga model yang terbentuk mampu meningkatkan hasil akurasi. Tidak semua *hyperparameter* perlu dilakukan proses *hypertuning*, seperti *learning rate*. Beberapa *hyperparameter* yang biasa diikutsertakan dalam *hypertuning* adalah

optimizer, *dropout rate*, dan lebar setiap layer (Li, dkk., 2018). *Optimizer* digunakan untuk memperbarui bobot secara iteratif yang didasarkan pada data *training*. Sedangkan *dropout rate* digunakan untuk menghindari *overfitting*.

Menurut Li, dkk. (2018), terdapat lima langkah dalam melakukan *hypertuning*, diantaranya:

1. Mengatur nilai *hyperparameter* berdasarkan pemahaman peneliti, algoritma model, dan masalah yang ingin diselesaikan.
2. Menganalisis pengaruh *hypertuning* dengan hasil dari semua percobaannya. Dengan dilakukannya beberapa percobaan, diharapkan mampu menjawab pertanyaan terkait *hyperparameter* mana yang paling berpengaruh dalam seluruh percobaan.
3. Memvalidasi hasil hipotesis langkah dua. Hal ini dilakukan dengan menelusuri setiap detail dari masing-masing percobaan yang sudah dilakukan.
4. Memutuskan apakah diperlukan lebih banyak lagi *hypertuning*.
5. *Riview* dan simpan *progress* jika tidak diperlukan lagi *hypertuning*.

Salah satu metode *hypertuning* yang dapat digunakan dalam klasifikasi adalah *GridsearchCV*. Menurut Ginting, dkk. (2022), *GridSearchCV* adalah salah satu metode yang digunakan dalam memilih kombinasi *hyperparameter* dengan cara menguji satu persatu kombinasi serta melakukan validasi pada setiap kombinasi yang terbentuk. Kombinasi terbaik yang dihasilkan dari *hypertuning GridsearchCV* ini diharapkan mampu menghasilkan performa model terbaik untuk melakukan prediksi atau klasifikasi.

2.6 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kinerja dari suatu model dalam mencapai hasil klasifikasi atau prediksi. Dalam melakukan evaluasi model, model digunakan untuk memprediksi hasil klasifikasi, baik itu *binary classification* maupun *multiclass classification* pada dataset *testing* dan kemudian bandingkan hasil klasifikasi yang diprediksi dengan hasil klasifikasi yang sebenarnya.

Diperlukan perhitungan *accuracy*, *sensitivity*, dan *specificity* dengan *confusion matrix* dalam mengevaluasi kinerja suatu model khususnya untuk kasus klasifikasi.

2.6.1 Confusion Matrix

Confusion matrix mempresentasikan hasil evaluasi model berupa prediksi dan kondisi aktual dari data dengan menggunakan tabel matrik. Umumnya jika dataset memiliki dua kelas, kelas pertama dianggap positif dan kelas kedua dianggap negatif. Berdasarkan *confusion matrix* yang telah diperoleh, dapat ditentukan nilai *Accuracy*, *Precision*, *Recall*, *Specificity*, dan *F1 Score* (Putra & Wibowo, 2020). *Confusion matrix* dapat ditunjukkan pada Tabel 1.

Tabel 1. *Confusion Matrix*

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	<i>True Negative (TN)</i>	<i>False Positive (FP)</i>
Aktual Positif	<i>False Negative (FN)</i>	<i>True Positive (TP)</i>

True Positive (TP) adalah jumlah data positif yang diklasifikasikan secara akurat. Sama halnya dengan TP, *True Negative (TN)* adalah jumlah data negatif yang diklasifikasikan secara akurat. Sedangkan *False Positive (FP)* adalah jumlah data aktual negatif yang diklasifikasikan sebagai positif dan *False Negative (FN)* merupakan jumlah data aktual positif yang diklasifikasikan sebagai negatif.

Accuracy adalah jumlah perbandingan data yang benar dengan jumlah keseluruhan data. Perhitungan akurasi ini dirumuskan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.15)$$

Precision digunakan untuk mengukur seberapa besar proporsi dari kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan hasil prediksi kelas positif.

$$Precision = \frac{TP}{TP + FP} \quad (2.16)$$

Recall digunakan untuk menunjukkan presentase kelas data positif yang berhasil diprediksi dengan benar dari keseluruhan data kelas positif.

$$Recall = \frac{TP}{TP + FN} \quad (2.17)$$

F1 *Score* adalah perbandingan rata-rata *precision* dan *recall* yang dibobotkan.

$$F1 \text{ Score} = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \quad (2.18)$$

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2022/2023 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari *website* Kaggle mengenai data *Electronic Health Record* (EHR) pasien dari hasil pemeriksaan laboratorium yang diambil dari salah satu rumah sakit swasta di Indonesia. Data berjumlah 4412 berberntuk tabel yang memiliki sebelas kolom yaitu *Haematrocit*, *Haemoglobins*, *Erythrocyte*, *Leucocyte*, *Thrombocyte*, *MCH*, *MCHC*, *MCV*, *Age*, *Sex*, dan *Source*. Data yang digunakan memiliki sepuluh atribut yaitu *Haematrocit*, *Haemoglobins*, *Erythrocyte*, *Leucocyte*, *Thrombocyte*, *MCH*, *MCHC*, *MCV*, *Age*, dan *Sex*. Sedangkan target dari data ini merupakan variabel *Source* yang memiliki dua kelas yaitu kelas *incare* dan *outcare*. Link data EHR ini yaitu <https://data.mendeley.com/datasets/7kv3rctx7m/1>. Data EHR yang digunakan pada penelitian ini ditunjukkan pada Tabel 2.

Tabel 2. Data *Electronic Health Record* (EHR)

No	<i>Haematro</i> <i>cit</i>	<i>Haemoglo</i> <i>bins</i>	<i>Erythro</i> <i>cyte</i>	<i>Leuco</i> <i>cyte</i>	<i>Thrombo</i> <i>cyte</i>	MCH	MCHC	MVC	Age	Sex	<i>Source</i>
1	35.1	11.8	4.65	6.3	310	25.4	33.6	75.5	1	F	out
2	43.5	14.8	5.39	12.7	334	27.5	34.0	80.7	1	F	out
3	33.5	11.3	4.74	13.2	305	23.8	33.7	70.7	1	F	out
:	:	:	:	:	:	:	:	:	:	:	:
44 10	33.2	11.2	3.47	7.2	235	32.3	32.3	95.7	93	F	out
44 11	31.5	10.4	3.15	9.1	187	33	33	100.0	98	F	in
44 12	33.5	10.9	3.44	5.8	275	31.7	32.5	97.4	99	F	out

Spesifikasi perangkat yang digunakan pada penelitian ini adalah laptop merek Acer model Aspire A514-51 KG. Berikut merupakan spesifikasi *hardware* laptop tersebut:

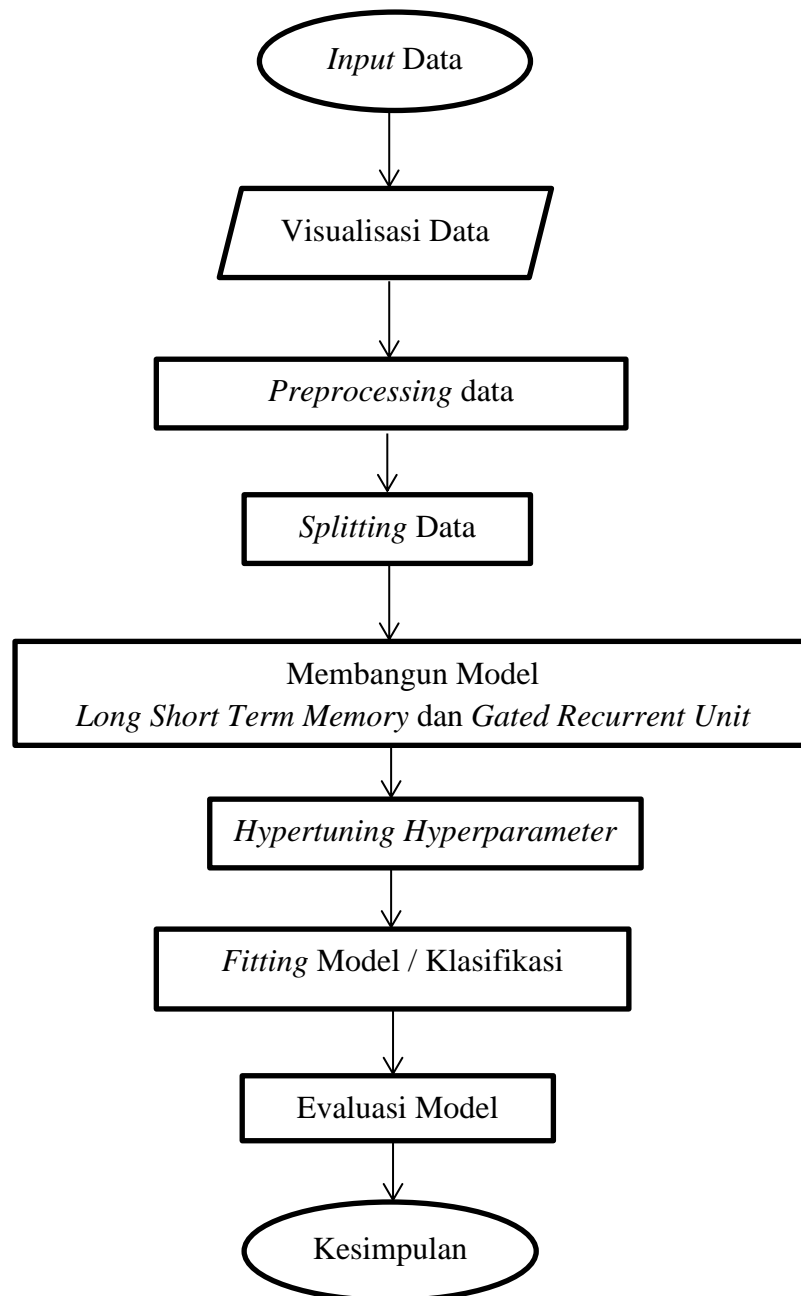
- Processor name : Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz 2.30GHz
- Processor speed : 2300MHz
- Memory : 4 GB DDR4

3.3 Metode Penelitian

Dalam penelitian ini, peneliti akan membandingkan dua model terbaik dari metode LSTM dan GRU untuk melakukan klasifikasi jenis pelayanan pasien rumah sakit berdasarkan data *Electronic Health Record* (EHR) dengan bantuan *software* Python yang didukung oleh Google Colab. Kemudian model yang dihasilkan akan dievaluasi berdasarkan *confusion matrix*. Adapun langkah-langkah yang dilakukan pada metode penelitian ini sebagai berikut:

1. Melakukan penginputan data *Electronic Health Record* (EHR) yang diperoleh melalui *website* Kaggle ke dalam Google Colab.
2. Melakukan eksplorasi data dengan cara visualisasi data dengan membuat grafik untuk melihat perbandingan jumlah pasien rawat inap dan rawat jalan.
3. Melakukan *preprocessing* data dengan mengecek keberadaan data hilang atau *missing value* dan mengecek keberadaan data yang berduplikat.
4. *Splitting* data yaitu membagi data menjadi data *training* dan data *testing*
5. Membangun model dengan menerapkan metode *Long Short Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU).
6. Melakukan *hypertuning hyperparameter* pada metode LSTM dan GRU.
7. *Fitting* model untuk mengklasifikasi jenis pelayanan pasien rumah sakit (pasien rawat inap dan rawat jalan) berdasarkan data EHR.
8. Evaluasi model menggunakan *confusion matrix*.
9. Memaparkan kesimpulan berdasarkan hasil yang diperoleh.

Gambar 6 berikut merupakan diagram alir untuk melihat proses penerapan metode LSTM dan GRU secara singkat menggunakan *software* Python yang didukung oleh Google Colab:



Gambar 6. Diagram Alir Metode LSTM dan GRU.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan sebagai berikut:

1. Pengujian ini menggunakan dua metode yaitu LSTM dan GRU yang masing-masing dilakukan satu model pengujian *hypertuning* untuk melakukan klasifikasi pelayanan rumah sakit.
2. Model LSTM memerlukan *running time* yang lebih lama daripada model GRU. Proses *hypertuning* untuk memperoleh model LSTM terbaik dibutuhkan *running time* selama 1 jam 4 menit 27 detik. Sedangkan proses *hypertuning* untuk memperoleh model GRU terbaik dibutuhkan *running time* selama 47 menit 24 detik.
3. Model LSTM terbaik yang diperoleh mempunyai *hyperparameter unit* sebesar 128 dan *batch size* 64. Dengan menggunakan epoch 200 dan *early stopping*, model LSTM tersebut berhenti di epoch 87 dan memiliki nilai akurasi sebesar 75.57%. Sedangkan model GRU terbaik yang diperoleh memiliki *hyperparameter unit* sebesar 64 dan *batch size* 32. Dengan menggunakan epoch 200 dan *early stopping*, model GRU tersebut berhenti di epoch 39 dan memiliki nilai akurasi sebesar 75.34%.
4. Berdasarkan hasil akurasi yang telah diperoleh dari kedua model, baik LSTM maupun GRU, nilai akurasi terbaik didapatkan dari model LSTM dengan nilai akurasi sebesar 75.57%.

DAFTAR PUSTAKA

- Francois, V., Henderson, P., Islam, R., Bellemere, M.G., & Pineau, J. 2018. An Introduction to Deep Reinforcement Learning. *Foundations and Trends in Machine Learning*. **11**(3-2): 219-354.
- Ginting, L.M., Sigiro, M., Lumbantoruan, G.Y., & Lumbangaol, J. 2022. Prediksi Indikator Perawatan Menggunakan Random Forest Classification dan Classification and Regression Trees. *Journal of Applied Technology and Informatics*. **1**(7): 43-48.
- Hamakonda, T.P. & Tairas, J.N.B. 2008. *Pengantar Klasifikasi Persepuluhan Dewey*. Ed. ke-5. PT BPK Gunung Mulia, Jakarta.
- Han, J., Kamber, M., & Pei, J. 2012. *Data Mining Concepts and Techniques*. Ed. Ke-3. Morgan Kaufmann, USA.
- Handari, B.D., Niman, I.M.S., Hasan, A., Purba, J.R.P., & Hertono, G.F. 2021. Comparison Of Elman Neural Network, Long Short Term Memory And Gated Recurrent Unit In Predicting Dengue Hemorrhagic Fever At DKI Jakarta. *Communication in Mathematical Biology and Neuroscience (CMBN)*. **1**(87): 1-25.
- Haris, N.M., Ardi, & Lawi, A. 2021. Klasifikasi Suara Detak Jantung Menggunakan Model Long-Short Term Memory Dan Gated Recurrent Unit. *Konferensi Nasional Ilmu Komputer (KONIK)*. **5**(1): 62-65.
- Hastomo, W., Aini, N., Karno, A.S.B., & Rere, L.M.R. 2022. Metode Pembelajaran Mesin untuk Memprediksi Emisi Manure Management. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*. **11**(2): 131-139.

- Heryadi, Y. & Irwansyah, E. *Deep Learning: Aplikasinya di Bidang Geospasial*. 2020. PT Artifisia Wahana Informa Teknologi, Jawa Barat.
- Li, T., Convertino, G., Wang, W., Most, H., Zajonc, T., & Tsai, Y.H. 2018. HyperTuner: Visual Analytics for Hypertuning by Professionals. *Conference: Workshop on Machine Learning from User Interaction for Visualization and Analytics (MLUI)*.
- Mavrogiorgou, A., Kiourtis, A., Touloupou, M., Kapassa, E., & Kyriazis, D. 2019. Internet of Medical Things (IoMT): Acquiring and Transforming Data into HL7 FHIR through 5g Network Slicing. *Emerging Science Journal*. **3**(2): 64–77.
- Manaswi, N. K. 2018. *Deep Learning with Applications Using Python*. Apress, New York.
- Pardede, J. & Raspati, M.F. 2021. Gated Recurrent Units dalam Mendeteksi Obstructive Sleep Apnea. *MIND (Multimedia Artificial Intelligent Networking Database) Journal*. **6**(2): 221-235.
- Putra, D. & Wibowo, A. 2020. Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Syntax Naïve Bayes, hlm. 84-92. Prosiding Seminar Nasional Riset Dan Information Science (SENARIS), Universitas Budi Luhur.
- Putra, R.B.R. & Hendry. 2022. Multivariate Time Series Forecasting pada Penjualan Barang Retail dengan Recurrent Neural Network. *Jurnal Inovtek Polbeng - Seri Informatika*. **7**(1): 71-82.
- Qiao, Z., Sun, N., Li, X., Xia, E., Zhao, S., & Qin, Y. 2018. Using Machine Learning Approaches for Emergency Room Visit Prediction Based on Electronic Health Record Data. *Stud. Health Technol. Inform.* 247: 111–115.
- Rosita, R., Pertiwi, D.A.A., & Khairunnisa, O.G. 2022. Prediction of hospital intensive patients using neural network algorithm *Journal of Soft Computing Exploration*. **3**(1): 8–11.

- Sadikin, M. & Nurhaida, I. 2021. Exploratory Study of Some Machine Learning Techniques to Classify the Patient Treatment. *International Journal of Advanced Computer Science and Applications*. **12**(2): 380-387.
- Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. 2018. Deep learning With Gated Recurrent Unit Networks for Financial Sequence Predictions. *Procedia Computer Science*. **131**: 895– 903.
- Sholeh, M.A. & Hidayat, R. 2022. Perbandingan Model LSTM dan GRU untuk Memprediksi Harga Minyak Goreng di Indonesia. *Edusaintek: Jurnal Pendidikan, Sains dan Teknologi*. **9**(3): 800-811.
- Singh, S., Pandey, S.K., Pawar, U., & Janghed, R.R. 2018. Classification of ECG Arrhythmia using Recurrent Neural Networks. *Procedia Computer Science*. **132**(1): 1290-1297.
- Sitepu, A.C. & Sigiyo, M. 2021. Analisis Fungsi Aktivasi Relu dan Sigmoid Menggunakan Optimier SGD dengan Representasi MSE pada Model. *JUTISAL (Jurnal Teknik Informatika Komputer Universal)*. **1**(1): 12-25.
- Tutuko, B., Nurmaini, S., Rachmatullah, M.N., Darmawahyuni, A., & Firdaus. 2020. A Deep Learning Approach to Integrate Medical Big Data for Improving Health Services in Indonesia. *Computer Engineering and Applications*. **9**(1): 17-28.
- Wicaksono, H., Liliana, & Tjondrowiguno, A.N. 2022. Pemodelan Lip Reading Bahasa Indonesia Berbasis Visem Menggunakan VGG16 serta Jaro-Winkler Similarity dan Bigram. *Jurnal Infra*. **10**(2): 1-7.