

## **BAB III**

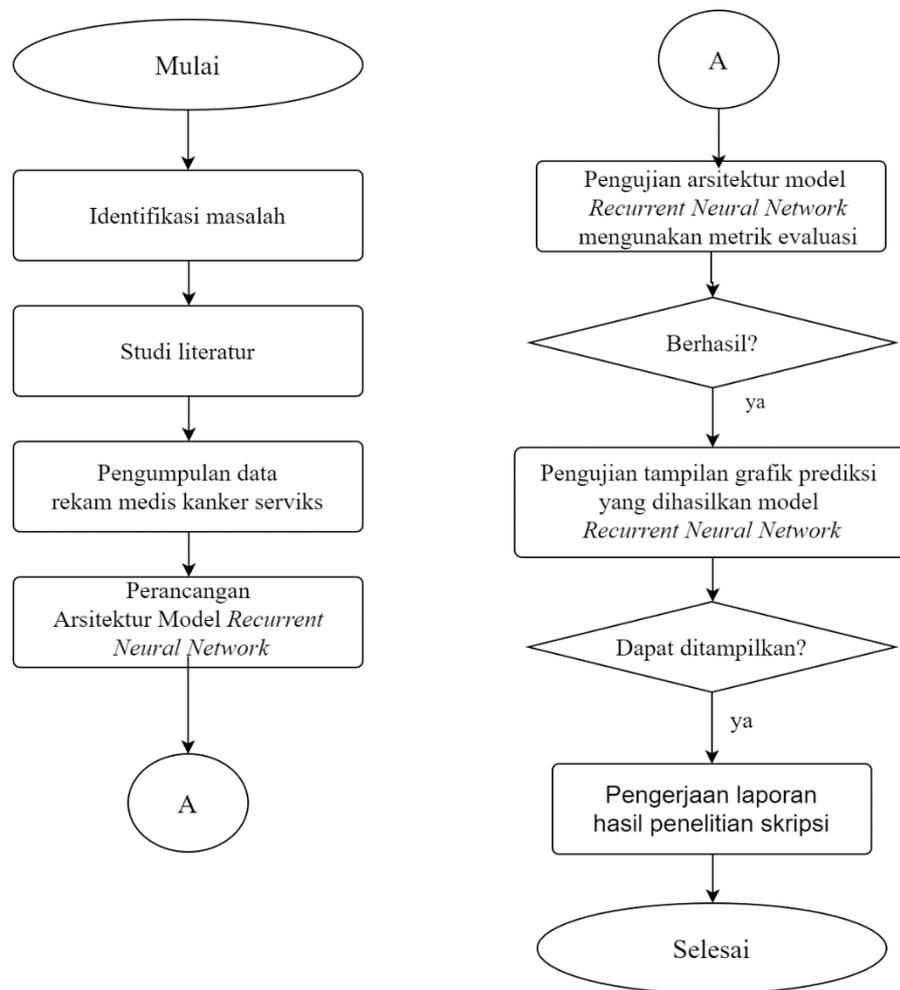
### **Metode Penelitian**

#### **3.1 Jenis Penelitian**

Jenis penelitian yang diadopsi adalah penelitian ini adalah *Research & Development* (R&D) menggunakan Pendekatan pengerjaan *Machine learning Workflow* yang terdiri dari 5 tahapan. Hal ini dikarenakan penelitian ini terkait dengan perancangan suatu sistem dan memiliki sifat yang khusus terinci dan statis (Muttaqin & Firmansyah, 2022). Tujuan dari penelitian ini adalah membuat model RNN yang dapat mendeteksi kanker serviks untuk mempersingkat waktu dokter untuk mengidentifikasi penyakit kanker dengan *menginputkan* parameter dalam rekam medis kanker serviks kedalam model RNN. Model ini diharapkan dapat membantu mempersingkat waktu dokter untuk membantu untuk melakukan diagnosa kanker kepada pasien.

#### **3.2 Alur Penelitian**

Dalam melakukan suatu penelitian, maka diperlukan suatu langkah atau alur agar penelitian dapat dilakukan dengan lancar dan terstruktur. Dalam penelitian ini *flowchart* yang digunakan untuk menjelaskan tahapan penelitian adalah sebagai pada gambar 3.1 berikut :



Gambar 3. 1 *Flowchart* penelitian

Untuk penjelasan dari gambar 3.1 lebih jelasnya terdapat di bawah ini :

### 3.2.1 Identifikasi masalah

Tahap penelitian dimulai dengan identifikasi masalah, identifikasi masalah merupakan suatu proses untuk menentukan poin masalah yang sudah dijelaskan pada latar belakang pada bab 1. Dalam penelitian ini sudah ditentukan beberapa point masalah yang harus diselesaikan yaitu perancangan arsitektur model RNN untuk mendeteksi kanker serviks pada pasien CMI Hospital bandung, bagaimana penentuan *hyperparameter* yang digunakan pada arsitektur model RNN agar model

RNN bisa berjalan sebagaimana mestinya untuk melakukan deteksi dan prediksi keadaan pasien serta mendapatkan tingkat akurasi yang dihasilkan oleh model RNN berdasarkan data input rekam medis kanker serviks CMI Hospital Bandung.

### 3.2.2 Studi Literatur

Langkah selanjutnya adalah melakukan studi literatur. Studi literatur merupakan proses pengumpulan informasi dari berbagai sumber, baik dari sumber media cetak atau dari media *online*. Dalam penelitian ini, studi literatur banyak diambil dari Internet seperti dari *google scholar*, *Elsevier*, *IEEE* serta *Semantic Scholar* yang beberapa laman ini merupakan laman untuk mendapatkan *paper*, jurnal serta buku yang sudah ada dalam bentuk media *online*. Adapun kata kunci utama yang dicari agar lebih terfokus pada penelitian yaitu *Deep learning*, *Recurrent neural network* (RNN), *Cervical Cancer*, untuk abstrak dalam bahasa Inggris dan *Deep learning*, *Recurrent Neural Network* dan kanker serviks untuk abstrak dalam bahasa Indonesia.

Dalam melakukan proses studi literatur ini dilakukan penentuan poin gagasan – gagasan terbaru dalam bidang *deep learning* dan RNN serta penggunaannya dalam lingkup kanker serviks. Selain itu untuk memahami perkembangan terbaru serta mengidentifikasi kekosongan pengetahuan yang dapat diisi melalui penelitian ini. Nantinya hasil studi literatur akan menjadi landasan yang penting untuk merancang metodologi penelitian untuk membangun argumen ilmiah yang kuat serta dasar teori yang kuat dalam pembuatan Model RNN yang baik.

### 3.2.3 Pengumpulan data rekam medis

Selanjutnya adalah mendapatkan data rekam medis kanker serviks, untuk memulai pengumpulan rekam medis, diperlukan surat rekomendasi penelitian dari universitas yang didapat dari bagian akademik sebagai surat resmi untuk izin melakukan penelitian pada CMI Hospital Bandung. Surat rekomendasi penelitian dari universitas dari bagian akademik dalam penelitian ini tidak diperlukan dikarenakan peneliti memiliki rekanan seorang dokter senior yang bertanggung jawab terhadap semua pasien kanker di CMI Hospital Bandung. Peneliti diminta presentasi secara singkat bagaimana Arsitektur RNN kepada dokter secara daring

Fairuz Fernanda Hermawan, 2023

*Desain Sistem Deteksi Kanker Serviks Menggunakan Algoritma Recurrent Neural Network pada Metode Deep Learning*

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

via *Google-Meet* yang dipilih sebagai penilai untuk meloloskan atau tidak proposal penelitian yang diajukan oleh penulis. Setelah mendapat persetujuan dari dokter penanggung jawab pasien kanker CMI Hospital Bandung, sekitar seminggu kemudian penulis dipanggil oleh dokter yang bersangkutan ke CMI Hospital Bandung untuk mengambil data rekam medis kanker serviks dalam bentuk *hardcopy* sebanyak 600 lembar.

### 3.2.4 Perancangan Arsitektur Model RNN

Selanjutnya adalah melakukan perancangan arsitektur model RNN, hal ini meliputi pencarian *source code* sebagai acuan kerangka model, melakukan proses coding mulai dari penginputan *library*, penginputan data rekam medis dalam format CSV, penentuan *hyperparameter*, penentuan *input layer*, *hidden layer* dan *output layer* sampai proses pelatihan model RNN itu sendiri yang semuanya dieksekusi dalam Bahasa pemrograman *python*.

### 3.2.5 Pengujian Arsitektur Model RNN menggunakan metrik evaluasi

Selanjutnya adalah pengujian arsitektur model RNN menggunakan metrik evaluasi, dalam penelitian ini metrik evaluasi yang digunakan adalah *Mean Squared error* (MSE), MSE ini secara garis besar adalah metrik untuk menghitung seberapa akurat dalam memprediksi nilai angka, MSE menghitung jumlah rata rata dari selisih kuadrat dari nilai prediksi yang dihasilkan model RNN dan nilai aktual yang diambil dari data rekam medis pada CSV sebelumnya untuk batas MSE yang ditentukan dalam penelitian ini adalah dibawah 1. Alasannya adalah jika nilai MSE model RNN diatas satu maka nilai prediksi jelas akan berbeda jauh dengan data aktual.

Adapun persamaan dari MSE adalah sebagai berikut :

$$MSE = \sum \frac{(Y' - Y)^2}{n} \quad 3.1$$

Metrik kedua yang digunakan adalah *Root Mean Squared Error* atau RMSE. *Root Mean Squared Error* (RMSE) adalah salah satu metrik evaluasi yang umum digunakan dalam model regresi, termasuk dalam model RNN. RMSE mengukur sejauh mana perbedaan antara nilai prediksi yang dihasilkan oleh model dengan nilai aktual yang terdapat dalam data.

Dalam konteks model RNN untuk data tipe regresi, seperti pada kasus deteksi kanker serviks, RMSE berfungsi dalam mengukur sejauh mana kesalahan prediksi model terhadap nilai aktual lebar luka *lesi* leher rahim pada pasien. RMSE menghitung rata-rata dari seluruh perbedaan kuadrat antara nilai prediksi dan nilai aktual. Kemudian, akar kuadrat dari rata-rata tersebut diambil untuk menghasilkan skala nilai yang sesuai dengan satuan data (dalam hal ini, milimeter).

Gunanya, RMSE memberikan informasi tentang tingkat kesalahan prediksi model secara keseluruhan. Semakin rendah nilai RMSE, semakin mendekati prediksi model dengan nilai aktual, dan semakin baik performa model dalam menghasilkan prediksi yang akurat. RMSE sangat bermanfaat dalam kasus seperti deteksi kanker serviks, di mana ketepatan dalam memprediksi lebar luka *lesi* leher rahim penting untuk memastikan hasil yang lebih akurat dan relevan dalam pengambilan keputusan klinis.

Adapun persamaan dari RMSE ini sendiri adalah sebagai berikut :

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}} \quad 3.2$$

Keterangan rumus tersebut adalah sebagai berikut :

$Y'$  = Nilai *Output* Prediksi yang dihasilkan model RNN

$Y$  = Nilai *Output* Aktual Yang dihasilkan model RNN

$n$  = Jumlah data yang ada

Metrik ketiga yang akan digunakan adalah *Mean Absoluter Error* (MAE). *Mean Absolute Error* (MAE) adalah metrik evaluasi yang mengukur rata-rata dari nilai absolut dari perbedaan antara nilai prediksi dan nilai aktual. Dalam konteks kasus kanker serviks dan model RNN, MAE akan memberikan informasi tentang seberapa besar kesalahan rata-rata dari prediksi model terhadap nilai lebar luka *lesi* leher rahim dalam skala absolut. Penggunaan MAE memiliki keuntungan dalam kasus ini karena ia mengukur kesalahan tanpa memperhitungkan arah perbedaan (positif atau negatif). Dalam konteks deteksi kanker serviks, penting untuk mengetahui seberapa besar kesalahan yang mungkin terjadi dalam memprediksi nilai lebar luka *lesi*, terlepas dari apakah prediksinya lebih besar atau lebih kecil dari nilai aktual. MAE memberikan informasi yang lebih intuitif dan mudah dimengerti bagi banyak orang, karena kesalahan diukur dalam satuan yang sama dengan data aktual.

Adapun rumus dari MAE adalah sebagai berikut :

$$MAE = \sum \frac{|Y' - Y|^2}{n} \quad 3.3$$

Keterangan rumus tersebut adalah sebagai berikut :

$Y'$  = Nilai *Output* data prediksi yang dihasilkan model RNN

$Y$  = Nilai *Output* data Aktual Yang dihasilkan model RNN

$n$  = Jumlah data yang ada

Sedangkan metrik evaluasi keempat yang digunakan adalah akurasi atau *accuracy*, akurasi merupakan metrik evaluasi yang digunakan untuk mengukur seberapa banyak prediksi yang benar dibandingkan dengan keseluruhan data.

Persamaan untuk akurasi ini adalah sebagai berikut :

$$Accuracy = \frac{(TP) + (TN)}{(TP) + (TN) + (FP) + (FN)} * 100\% \quad 3.4$$

Adapun keterangan dari persamaan ini yaitu :

- *True Positive (TP)* : kasus dimana model dapat memprediksi dengan benar sebagai contoh jika nilai data asli adalah positif maka data prediksi akan menghasilkan nilai positif juga.
- *True Negative (TN)* : kasus di mana model dapat memprediksi data aktual negatif dan menghasilkan prediksi data negatif dengan benar.
- *False Positive (FP)* : Kasus di mana model melakukan kesalahan di mana model menghasilkan nilai positif dimana seharusnya model menghasilkan nilai negatif.
- *False Negative (FN)* : kebalikan dari *False positive* dimana model menghasilkan nilai negatif dimana seharusnya model menghasilkan nilai positif.

### **3.2.6 Pengujian Tampilan grafik prediksi antara data prediksi dan data aktual model RNN**

Selanjutnya adalah pengujian model RNN apakah model dapat menampilkan grafik dari nilai yang sudah dihasilkan model RNN. Hal ini dilakukan agar pengamat dapat membaca dengan mudah antara nilai prediksi dan nilai aktual yang dihasilkan oleh model RNN.

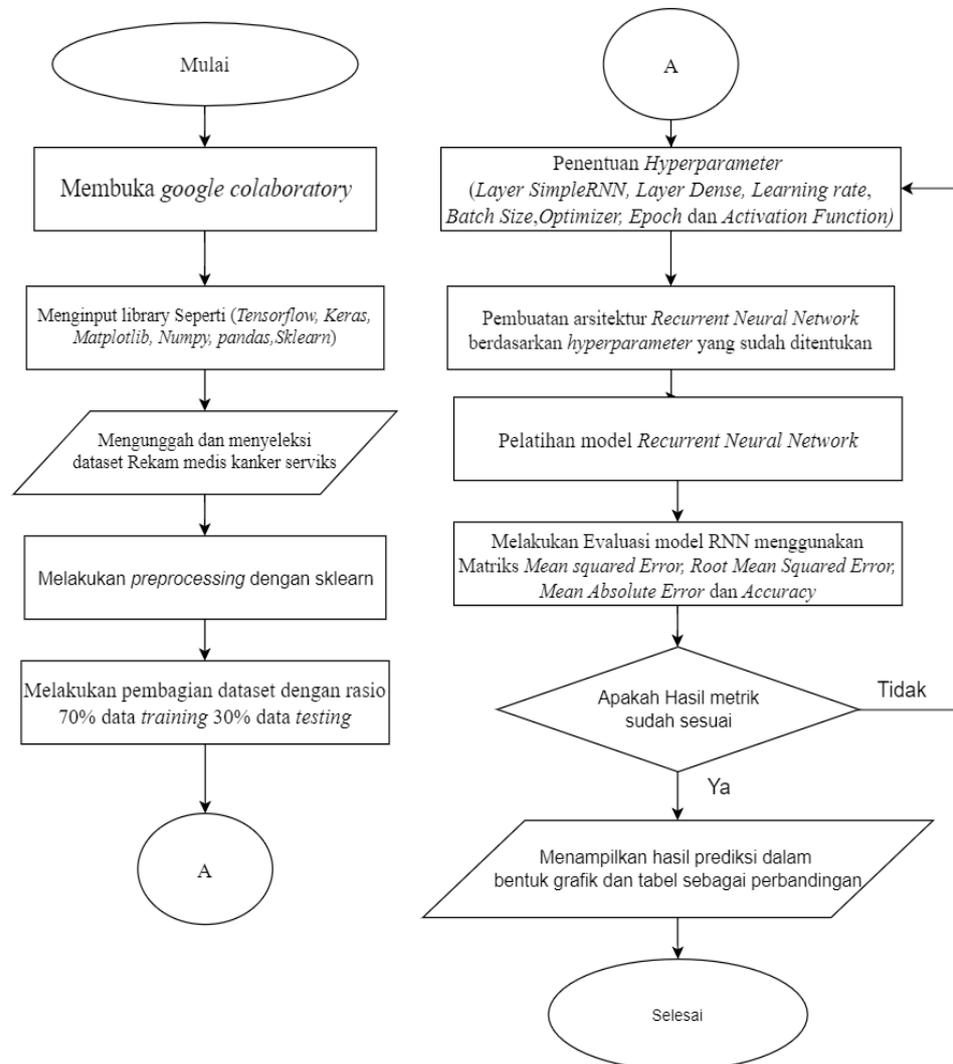
### **3.2.7 Pengerjaan Laporan Hasil Penelitian untuk keperluan skripsi**

Langka terakhir adalah pengerjaan laporan hasil penelitian yang dilakukan dari langkah awal sampai saat ini kedalam bentuk laporan hasil penelitian yang nanti akan dijelaskan pada Bab IV “Hasil dan Pembahasan” serta kesimpulan pada Bab V “Kesimpulan dan saran penelitian”.

## **3.3 Desain dan Flowchart Sistem**

*Flowchart* merupakan alur atau langkah langkah yang dibuatkan kedalam bentuk diagram. *Flowchart* biasanya digunakan untuk menggambarkan alur aliran kerja dari suatu program, prosedur atau proses dalam berbisnis. Diagram ini digambarkan menggunakan simbol dari berbagai jenis tindakan seperti *input/output*, keputusan dan penghubung. Berikut adalah *flowchart* perancangan

sistem untuk mendeteksi kanker serviks menggunakan model RNN, seperti yang terlihat pada gambar 3.2



Gambar 3. 2 Flowchart desain Sistem Prediksi Kanker Serviks model RNN

Untuk penjelasan dari gambar 3.2 adalah sebagai berikut :

### 3.3.1 Membuka Google Colab

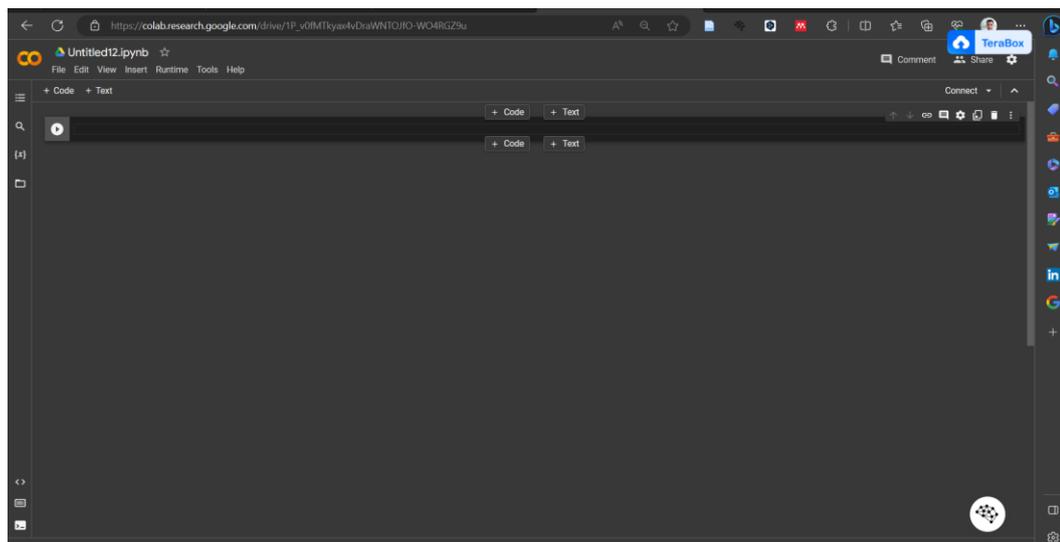
Langkah pertama dalam membuat model RNN adalah membuka laman *Google colaboratory* pada link <https://colab.research.google.com/>. kemudian setelah masuk ke laman *Google Colab*, membuat *new notebook* dan memberi nama untuk *file* nya

Fairuz Fernanda Hermawan, 2023

*Desain Sistem Deteksi Kanker Serviks Menggunakan Algoritma Recurrent Neural Network pada Metode Deep Learning*

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

dengan nama yang sesuai. Setelah melakukan semua dengan benar maka akan muncul laman seperti gambar 3.3 berikut :



Gambar 3. 3 *Environment* awal pada *google colab*

### 3.3.2 Menginput *library*

Langkah selanjutnya adalah *mengimport library*. *library* merupakan kumpulan kode yang berisi fungsi, metode dan kelas yang sudah dituliskan sebelumnya oleh pihak pengembang bersangkutan. Tujuan dari penggunaan *library* ini adalah mempercepat dan mempermudah proses pembuatan model RNN. Mengunduh dan membaca *dataset* rekam medis kanker serviks.

Langkah selanjutnya adalah mengunggah data rekam medis dengan format berbentuk csv kedalam *Google Colab* kemudian memanggil data rekam medis yang sudah diubah kedalam bentuk CSV kedalam kode.

### 3.3.3 Memulai *Preprocessing* dengan *sklearn*

Langkah selanjutnya adalah melakukan *preprocessing* menggunakan *library sklearn*, dalam hal ini *preprocessing* dilakukan untuk menentukan nilai bawah (*min*) dan nilai atas (*max*) agar dapat ditentukan nilai bawah dan nilai atas dari data *Lesi* yang sudah diinput tadi.

### 3.3.4 Melakukan proses *data splitting*

Langkah selanjutnya adalah proses pembagian *dataset*, agar proses training model RNN dan prediksi yang dihasilkan akurat, maka data csv yang sudah diinput perlu dibagi menjadi dua yaitu data *training data* dan *data testing* dengan perbandingan 70% *data training* dan 30% *data testing*. agar model RNN dapat melakukan pengenalan pola serta membuat prediksi dengan baik namun tetap memperhatikan dari data rekam medis kanker serviks (Abraham et al., 2021; Rukhsar et al., 2022).

### 3.3.5 Penentuan *Hyperparameter*

Langkah selanjutnya adalah menentukan batasan untuk model RNN. Hal ini disebut dengan *hyperparameter*. Selain itu *hyperparameter* akan bekerja dan berkaitan pada proses pembuatan arsitektur RNN dan proses *training* model RNN, beberapa *hyperparameter* yang digunakan dalam model RNN pada penelitian ini adalah sebagai berikut :

#### 1. *SimpleRNN layer*

Jumlah *layer* dalam *SimpleRNN* akan mempengaruhi sejauh mana model mampu menggali informasi dari data berurutan. Jumlah *layer* ini akan menentukan tingkat representasi dan kompleksitas yang dapat dicapai oleh model RNN. Dalam penelitian ini digunakan jumlah layer pada *SimpleRNN* sebanyak 64 *neuron*.

#### 2. *Dense layer*

*Dense Layer* yang mengikuti *SimpleRNN* juga memiliki jumlah *layer* yang dapat diatur. Hal ini memungkinkan model RNN untuk melakukan ekstraksi fitur lebih lanjut dari representasi yang diberikan oleh *layer* sebelumnya. Dalam penelitian ini digunakan jumlah layer dense untuk mengolah informasi data rekam medis sebanyak 64 *neuron*. Dengan *output* layernya digunakan sebanyak 1 *neuron*.

#### 3. *Seq\_length*

*Seq\_length* menentukan berapa banyak time step dalam sekuensi yang akan dipertimbangkan oleh model. Ini memungkinkan model untuk memahami konteks temporal dari data pasien. *Seq\_length* yang digunakan pada penelitian ini adalah sebanyak 30 time step. 30 *time step* diambil karena jika terlalu tinggi maka akan

memakan memori dan daya komputasi lebih besar, serta menyebabkan *vanishing gradient*. Namun, jika diambil terlalu rendah maka akan kehilangan informasi dalam mengenali pola dan terjadi *overfitting* (model terlalu cocok pada data training) dan *underfitting* (model tidak cukup kompleks untuk memahami data).

#### 4. *Optimizer*

Pilihan *optimizer* mempengaruhi bagaimana model diperbarui selama proses pelatihan. Dalam penelitian ini, telah digunakan *optimizer* tertentu yang mengatur laju pembelajaran model. Dalam penelitian ini *optimizer* yang digunakan adalah *Root mean Squared Propagation* (RMSProp). Tujuan digunakan RMSprop adalah karena model RNN ini ditugaskan untuk memprediksi tugas regresi, lalu untuk mendapatkan hasil prediksi yang baik namun dengan kesalahan yang kecil pada saat diuji menggunakan metrik evaluasi MSE, RMSE dan MAE maka digunakanlah RMSprop.

#### 5. *Learning\_rate*

*Learning rate* mengontrol seberapa besar perubahan yang diterapkan pada bobot model setiap kali model diperbarui. Nilai learning rate yang sesuai memainkan peran penting dalam stabilitas dan kecepatan konvergensi pelatihan. Dalam penelitian ini digunakan nilai *learning rate* adalah 0,001 agar model RNN dapat mengatur parameter model yang sudah ditentukan selama menjalani proses *Model Training*. Alasan mengambil nilai 0.001 adalah merupakan nilai yang umum digunakan pada pelatihan *machine learning dan deep learning*.

#### 6. *Batch\_size*

*Batch size* adalah jumlah sampel yang diperlakukan bersamaan sebelum bobot model diperbarui. *Batch size* yang tepat dapat mempengaruhi kinerja dan efisiensi pelatihan *batch size* yang digunakan dalam penelitian ini adalah 5. Alasannya adalah agar model dapat menghitung 5 data sampel setiap melakukan satu kali iterasi untuk menghitung perubahan bobotnya.

#### 7. *Epoch*

Jumlah *epoch* menentukan berapa kali model akan melewati seluruh *dataset* pelatihan. Jumlah *epoch* yang ada penting untuk mencapai hasil yang akurat dan

konvergensi pelatihan. Dalam penelitian ini *epoch* yang digunakan adalah 100, alasannya adalah agar model dapat mempelajari pola dari data rekam medis dengan baik.

### **3.3.6 Pembuatan Arsitektur RNN**

Setelah penentuan beberapa *hyperparameter* yang akan digunakan, maka langkah selanjutnya adalah membuat kerangka model RNN. Secara umum model RNN memiliki 3 *layer* yaitu *input layer*, *hidden layer* serta *output layer*.

### **3.3.7 Melatih Model RNN**

Langkah selanjutnya adalah melakukan *training* Model RNN, tujuan dilakukan training agar model RNN dapat melakukan prediksi, pembuatan pola prediksi pada grafik yang akan ditunjukkan nantinya berdasarkan data aktual CSV kanker serviks.

### **3.3.8 Melakukan evaluasi model RNN (MSE, RMSE, MAE dan akurasi)**

Setelah melakukan pelatihan yang memakan waktu, maka langkah selanjutnya adalah mengevaluasi model RNN menggunakan metrik evaluasi MSE, RMSE, MAE dan metrik akurasi. Untuk hasilnya akan dibahas pada bab IV “Temuan dan Pembahasan”.

### **3.3.9 Menampilkan hasil Prediksi dalam bentuk grafik dan tabel**

Langkah terakhir adalah melakukan perbandingan antara hasil prediksi (ditunjukkan pada data prediksi) dengan data aktual (data yang diambil langsung dari csv) sebagai perbandingan. Adapun fungsi menampilkan grafik dan tabel adalah agar mudah melakukan pengidentifikasian bagi orang yang awam.

## **3.4 Populasi dan sampel data**

Penelitian ini berfokus pada masalah yang dihadapi oleh populasi pasien yang terkena kanker serviks, dengan kriteria berupa pasien kanker serviks dalam stadium 1 dengan fokus lebar *lesi* antara 3 hingga 5 mm. Populasi ini menjadi fokus utama penelitian karena kanker serviks merupakan salah satu jenis kanker yang signifikan dalam hal angka kejadian dan dampak kesehatan global.

Sedangkan sampel data yang digunakan dalam penelitian ini adalah lebar *lesi* luka leher rahim dengan periode mulai hari 1 sampai 600 untuk dijadikan data input yang diolah pada model RNN.

### 3.5 Alat Penelitian

Suatu penelitian penting untuk menggunakan alat dan bahan yang tepat untuk menunjang berlangsungnya penelitian. Alat yang digunakan untuk penelitian ini adalah Laptop yang terdiri dari perangkat keras dan perangkat lunak. Sedangkan bahan yang dipersiapkan adalah data rekam medis sebagai input dan hasil prediksi sebagai *output* :

#### 3.5.1 Alat Penelitian

Alat yang digunakan dalam penelitian ini bukan syarat minimum untuk membuat dan menjalankan model RNN, berikut merupakan alat yang digunakan dalam penelitian ini :

1. Perangkat keras (Laptop)
  - a. CPU : *Ryzen 7 5700U @ 1,8Ghz (16 cpus)*
  - b. GPU : *AMD Radeon Graphic Processor*
  - c. RAM : 8192 Mb
  - d. SSD : 512 GB
  - e. *Mouse dan Keyboard*
2. Perangkat Lunak
  - a. *Operating System : Windows 11 Home Version 10.0.2261 build 22621*
  - b. *Google Colab*
  - c. *Versi Python 3.12*
  - d. *Versi Tensorflow 2.12.0*
  - e. *Versi NumPy 1.22.4*
  - f. *Versi Pandas 1.5.3*