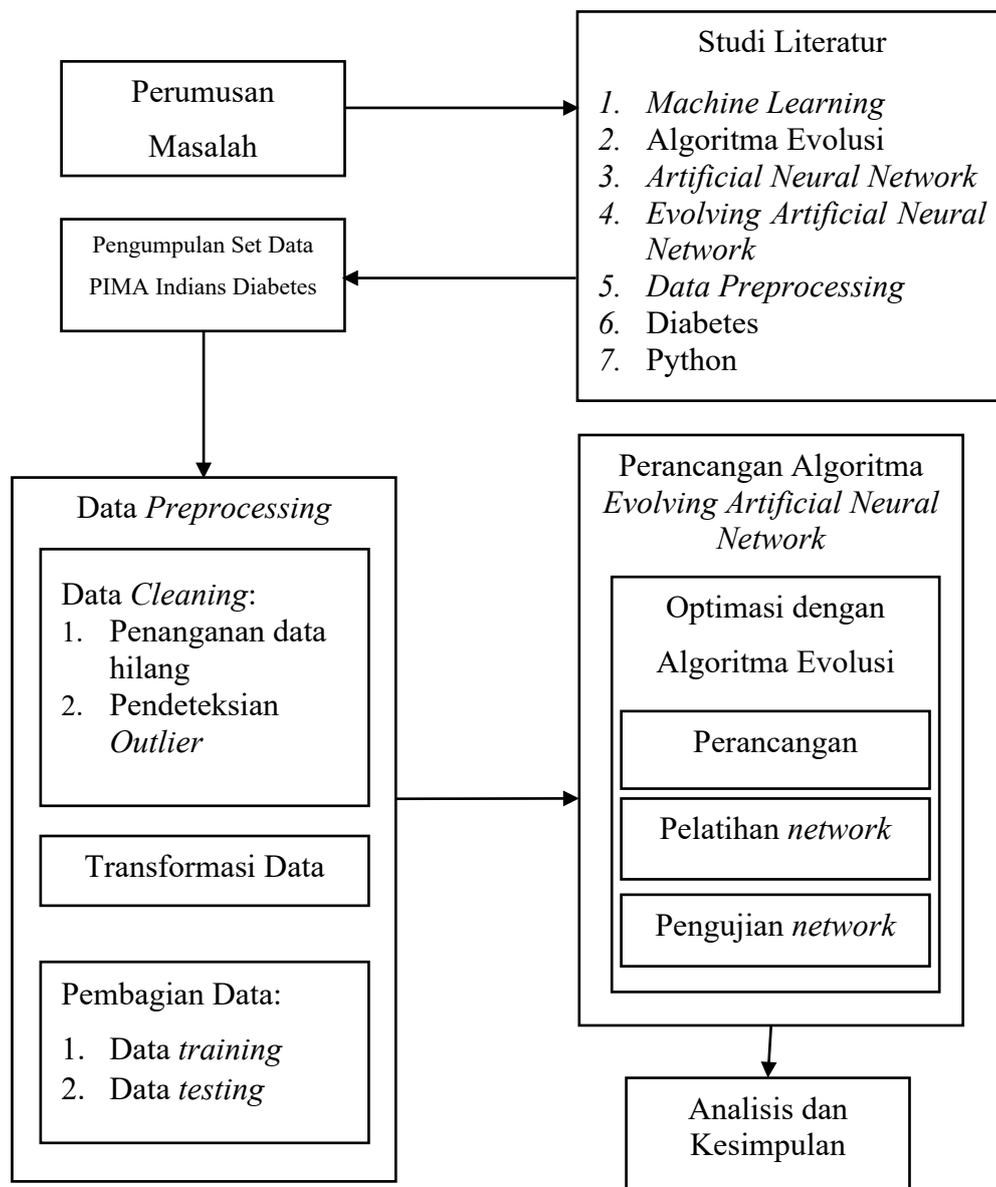


## BAB III METODOLOGI PENELITIAN

### 3.1 Desain Penelitian

Desain penelitian merupakan kerangka kerja yang akan digunakan untuk melakukan penelitian. Pada bagian ini penulis akan memaparkan kerangka kerja mulai dari awal penelitian hingga selesai. Desain Penelitian dapat dilihat pada Gambar 3.1.



Gambar 3.1 Desain Penelitian

### 3.1.1 Perumusan Masalah

Perumusan masalah merupakan tahap awal dari penelitian. Proses yang terjadi di tahap ini dimulai dari mengidentifikasi latar belakang permasalahan yang akan dibahas, kemudian merumuskan permasalahan, lalu menentukan metode atau algoritma yang akan digunakan untuk menyelesaikan masalah tersebut. Selain itu, perumusan masalah juga menghasilkan tujuan dan manfaat penelitian.

### 3.1.2 Studi Literatur

Studi literatur dilakukan terhadap topik-topik yang berkaitan dengan hasil dari perumusan masalah. Topik-topik tersebut antara lain adalah mengenai *machine learning*, algoritma evolusi, algoritma *artificial neural network*, algoritma *evolving artificial neural network*, *data preprocessing*, diabetes, dan bahasa pemrograman Python. Pada tahapan ini, penulis mencari studi literatur dari berbagai sumber referensi. Referensi-referensi tersebut antara lain berasal dari artikel jurnal, artikel konferensi, buku, dll.

### 3.1.3 Pengumpulan Set Data

Dalam penelitian ini data yang digunakan merupakan data sekunder. Dalam hal ini penulis tidak secara langsung mengambil data sendiri tetapi meneliti dan memanfaatkan data yang dihasilkan oleh pihak lain. Data yang akan digunakan pada penelitian ini diunduh dari situs Kaggle. Detail mengenai set data yang digunakan akan disebutkan pada subbab bahan penelitian.

### 3.1.4 Data Preprocessing

*Data preprocessing* merupakan tahapan untuk melihat tingkat kualitas dari data yang diperoleh. Permasalahan yang ada pada data antara lain adalah mungkin data yang diperoleh terlalu banyak, terlalu sedikit, ada data yang hilang, ada data aktual di luar hasil yang diinginkan (*outlier*), dan data yang diperoleh dalam skala yang berbeda. Hal ini dilakukan agar hasil prediksi yang diinginkan memiliki tingkat akurasi dan presisi yang tinggi.

Kemudian data juga akan ditransformasi, tahapan ini akan dilakukan terhadap data untuk melakukan standarisasi. Tujuannya adalah untuk mengubah nilai

numerik dalam setiap variabel sehingga setiap variabel memiliki nilai rata-rata dan variansi bernilai 0. Selain itu, standarisasi data ini juga bertujuan untuk menyusutkan besarnya nilai suatu variabel dan mengubahnya menjadi skala yang lebih proporsional.

Terakhir, data akan dibagi menjadi data *training* dan data *testing*. Data *training* akan digunakan pada tahapan *training*, sedangkan data *testing* akan digunakan pada tahap *testing*. Tujuan membagi data ke dalam set *training* dan *testing* adalah untuk menghindari *overfitting* dan untuk menyediakan sumber data yang tidak bias ketika mengevaluasi kinerja model.

### 3.1.5 Perancangan Algoritma

Metode yang digunakan penulis dalam mengimplementasikan perancangan algoritma dalam skripsi ini adalah dengan menggunakan metode air terjun atau *waterfall*. Terdapat lima tahapan dalam metode ini, yaitu *requirements analysis and definition, system and software design, implementation and unit testing, integration and system testing, dan operation and maintenance*. Penjelasan mengenai metode ini mengacu pada subbab 2.2.8 Metode *Waterfall*.

Perancangan algoritma dimulai dengan mempersiapkan dan membangun algoritma. Setelah algoritma telah dibuat, tahap selanjutnya adalah melatih *network* dari ANN dalam algoritma evolusi. Dalam melatih *network*, terdapat berbagai parameter yang harus dipersiapkan. Setelah seluruh parameter dipersiapkan, algoritma dapat dijalankan untuk melakukan pelatihan. Dalam melakukan pelatihan, seluruh gen *network* yang ada di dalam populasi akan melakukan pelatihan (*training*) dan pengujian (*testing*) di setiap evolusi. Gen terbaik yang didapatkan adalah gen yang memiliki akurasi tertinggi berdasarkan hasil dari tahapan *testing*. Selain itu, penulis juga akan melakukan perancangan algoritma *artificial neural network* menggunakan *software library* Scikit-learn dan Keras untuk dianalisis dan dibandingkan performanya terhadap algoritma *native* yang telah dirancang oleh penulis.

Lalu, performa dari model dapat diukur menggunakan *confusion matrix*. Data dari *confusion matrix* ini kemudian dapat menghasilkan nilai akurasi, *precision*,

*recall*, dan *f1-score*. Penjelasan mengenai metode *confusion matrix* ini merujuk pada penjelasan tahapan pengukuran performa dalam subbab 2.2.3 *Artificial Neural Network* (ANN)

### 3.1.6 Analisis dan Kesimpulan

Data performa dari seluruh model terbaik yang sudah dihasilkan oleh algoritma EANN melalui berbagai skenario eksperimen akan dicatat dan dianalisis. Data tersebut akan dianalisis mengenai pengaruh satu sama lain antar parameter dengan waktu eksekusi yang diperlukan oleh algoritma. Kemudian data performa akurasi antar skenario juga akan dibandingkan. Mulai dari perbandingan akurasi dari algoritma EANN dengan menggunakan set data yang telah melalui tahap *data preprocessing* dengan data yang tidak melalui tahap *data preprocessing*. Serta membandingkan akurasi dari algoritma yang menggunakan *software library*.

## 3.2 Alat dan Bahan Penelitian

Dalam pembangunan perangkat lunak prediksi diagnosis penyakit diabetes dengan metode *Artificial Neural Network* membutuhkan beberapa alat serta perangkat lunak pendukung. Berikut ini merupakan data rinci dari alat dan bahan yang digunakan untuk melakukan penelitian.

### 3.2.1 Alat Penelitian

- a. Perangkat Keras yaitu sebuah laptop ASUS X450JF dengan spesifikasi:
  - Processor Intel(R) Core(TM) i7-4700HQ CPU @ 2.40GHz, 2401Mhz, 4 Cores, 8 Logical Processors
  - GPU NVIDIA GeForce GT 745M 2GB
  - RAM 8 GB DDR3 1600 Mhz
  - Hard Drive 1 TB
- b. Perangkat Lunak
  - Sistem Operasi Windows 10
  - Python 3.7
  - Anaconda

- Jupyter Notebook
- Google Chrome

### 3.2.2 Bahan Penelitian

Bahan yang diperlukan untuk melakukan penelitian yaitu data diagnosis pasien penderita diabetes. Data ini diperoleh dari <https://www.kaggle.com/uciml/pima-indians-diabetes-database/download>, data tersebut merupakan milik National Institute of Diabetes and Digestive and Kidney Diseases dan lisensi dari University of California, Irvine.