

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang Penelitian**

Diabetes Melitus merupakan salah satu masalah dalam dunia kesehatan yang dapat menyerang siapa saja, mulai dari anak-anak, remaja, hingga orang dewasa. Oleh karena itu, Diabetes merupakan salah satu Penyakit Tidak Menular (PTM) yang saat ini telah menjadi ancaman serius bagi kesehatan global. Dikutip dari data World Health Organization (WHO), 70% dari total kematian di dunia disebabkan oleh penyakit tidak menular. Bahkan hingga tahun 2016, diabetes menempati peringkat ketujuh penyebab kematian di seluruh dunia (World Health Organization, 2018). Sejak tahun 1980, jumlah penderita diabetes di seluruh dunia telah meningkat hampir dua kali lipat dari 4,7% menjadi 8,5% dari total populasi. International Diabetes Federation (IDF) bahkan memperkirakan jumlah penderita diabetes di seluruh dunia akan mencapai 700 juta orang pada tahun 2045 (Malanda dkk., 2019).

Fakta lain yang perlu diwaspadai adalah posisi diabetes sebagai salah satu *silent killer* di Indonesia mungkin saja naik bila penyakit ini tidak ditangani dengan baik. Di tahun 2016, persentase kematian akibat diabetes di Indonesia mencapai 6,7% dan merupakan yang tertinggi kedua setelah Sri Lanka. Angka ini cukup tinggi, karena 2 dari 3 penderita diabetes di Indonesia tidak mengetahui bahwa dirinya mengidap diabetes (World Health Organization, 2016). Data lain juga menunjukkan bahwa masalah utama diabetes di Indonesia ialah keterlambatan diagnosis, dimana 73,5% penderita diabetes tidak menyadari dirinya sakit (Kompas.com, 2012). *Riset Kesehatan Dasar* (2018) menyatakan, gangguan toleransi glukosa (*prediabetes*) mengalami kenaikan sebesar 10,2% di Indonesia. Sebuah penelitian yang dilakukan di Kanada terhadap para pengidap diabetes juga mendapatkan hasil bahwa 74,2% dari 7101 responden survei mengatakan bahwa mereka terlambat didiagnosis. Keterlambatan diagnosis diabetes mengacu pada ditemukannya komplikasi penyakit lain yang menyertai ketika dihasilkan diagnosis bahwa seseorang mengidap diabetes (Roche & Wang, 2014).

Selain membentuk berbagai kebijakan untuk pengendalian dan pencegahan diabetes, sudah terdapat beberapa penelitian juga yang dilakukan di dunia ataupun di Indonesia mengenai penyakit Diabetes. Penelitian-penelitian tersebut di antaranya melakukan penelitian untuk memprediksi diagnosis pasien berdasarkan data riwayat kondisi kesehatan serta hasil diagnosis pasien-pasien sebelumnya. Kumari dan Chitra melakukan prediksi untuk diagnosis diabetes menggunakan algoritma *Support Vector Machine* (SVM), salah satu metode pada *machine learning*, yaitu *supervised learning* (Kumari & Chitra, 2013). Kemudian terdapat pula penelitian yang menggunakan *Modified Radial Basis Functional Neural Networks* (MRBF), metode ini juga merupakan salah satu metode *supervised learning*. Kedua penelitian ini menggunakan set data diabetes dari Pima Indians. Penelitian yang menggunakan MRBF ini menunjukkan hasil yang lebih baik daripada metode *Support Vector Machine* (Magudeeswaran, 2013). Penjelasan selengkapnya mengenai penelitian terdahulu terdapat pada subbab 2.1.

Salah satu dari berbagai macam algoritma yang dikenal juga dalam *machine learning* adalah algoritma *Artificial Neural Network* (ANN). Secara sederhana, ANN merupakan sebuah model komputasi yang terinspirasi dari ilmu biologi, yang terdiri dari elemen pemrosesan yang disebut neuron dan koneksi/jaringan di antara mereka yang disebut dengan koefisien atau bobot (Shanmuganathan & Samarasinghe, 2016). Beberapa poin keunggulan dari ANN dalam konteks untuk memprediksi diagnosis medis adalah tidak terlalu membutuhkan proses *training* dengan keilmuan statistika untuk dikembangkan, kemampuan untuk secara implisit mendeteksi hubungan non-linier yang kompleks antara variabel dependen dan independen, kemampuan untuk mendeteksi semua interaksi yang mungkin antara variabel prediktor, dan dapat dikembangkan dengan beberapa algoritma untuk proses *training* (Tu, 1996). Salah satu penggunaan ANN di dunia medis yaitu digunakan untuk memprediksi diagnosis penyakit nefritis dan penyakit hati dengan menggunakan set data dari UCI. Setelah dilakukan pengujian terhadap *testing set*, hasilnya menunjukkan performa yang sangat baik. *Network* dapat mengklasifikasikan dengan akurasi sebesar 95% (Al-Shayea, 2011).

Namun, algoritma ANN memiliki permasalahan utama dalam hal penentuan arsitektur yang tepat. Maksud dari arsitektur di sini adalah struktur dan pembobotan antar *node* di dalam ANN. Permasalahan ini dapat dipandang sebagai permasalahan optimasi, dimana terdapat banyak sekali kemungkinan arsitektur yang bisa terjadi (Pratomo, Dayawati, & Suyanto, 2009). Permasalahan lainnya adalah mengenai keakuratan algoritma ANN dapat sangat terdegradasi apabila masih terdapat *noisy* pada set data atau *irrelevant features* (Karegowda dkk., 2012). Maka dari itu, untuk meningkatkan akurasi dari prediksi akan ada tahapan untuk melakukan *preprocessing* pada set data dan juga tahapan untuk menggunakan algoritma evolusi, karena algoritma ini sangat cocok untuk diterapkan pada kasus optimasi (Pratomo dkk., 2009).

Terdapat tiga tingkat bagian pada penggabungan algoritma evolusi dan ANN, yaitu : *connection weight*; *architecture*; dan *learning rules*. Evolusi arsitektur memungkinkan ANN untuk menyesuaikan topologi mereka dengan tugas yang berbeda tanpa campur tangan manusia dan dengan demikian memberikan pendekatan untuk menghasilkan desain arsitektur ANN secara otomatis karena bobot dan struktur koneksi ANN dapat dikembangkan. Evolusi *learning rules* dapat dianggap sebagai proses "*learning to learn*" pada ANN, dimana adaptasi *learning rules* dicapai melalui evolusi. Ini juga dapat dianggap sebagai *adaptive process of automatic discovery of novel learning rule* (Yao, 1999).

Dengan berbagai paparan di atas, mulai dari penelitian sejenis dengan berbagai algoritma yang digunakan, kemudian pemaparan mengenai algoritma ANN beserta kelebihan dan kekurangan serta salah satu cara untuk mengoptimasi algoritma ANN. Maka dalam penelitian ini penulis akan mengimplementasikan algoritma *Evolving Artificial Neural Network* (EANN) untuk memprediksi diagnosis pasien. Dengan harapan penelitian ini dapat menghasilkan akurasi yang lebih tinggi dalam memprediksi diagnosis pasien dalam hal penyakit diabetes.

Set data yang akan digunakan dalam skripsi ini menggunakan data sekunder yang berasal dari set data Pima Indians Diabetes yang disediakan oleh "National Institute of Diabetes and Digestive and Kidney Diseases" dan dikelola oleh Kaggle.

Suku Pima Indian adalah sekelompok penduduk asli Amerika yang tinggal di Arizona. Mereka adalah kelompok orang yang sering kali diteliti atau dipelajari karena kecenderungan genetik mereka yang sangat tinggi terhadap diabetes. Dipercayai bahwa orang Pima Indian membawa gen yang memungkinkan mereka untuk bertahan hidup dalam menghadapi kelaparan yang lama. Gen ini memungkinkan orang Pima Indian untuk menyimpan glukosa dan karbohidrat di dalam tubuh mereka apa pun yang mereka makan, yang secara genetik hal ini menguntungkan untuk mereka yang berada di lingkungan atau tempat dimana kelaparan biasa terjadi (Loy, 2019). Set data Pima Indians Diabetes ini pernah digunakan di beberapa penelitian serta buku. Dalam situs Kaggle, set data ini juga termasuk set data yang banyak digunakan para pengguna. Oleh karena itu, skripsi ini akan menggunakan set data ini sebagai bahan penelitian karena set data ini memiliki kredibilitas yang tinggi untuk digunakan dalam penelitian.

## 1.2 Rumusan Masalah Penelitian

Rumusan masalah yang menjadi fokus pada pembahasan dalam skripsi ini adalah:

1. Bagaimana mengolah dan memproses data riwayat diagnosis pasien dari set data Pima Indians Diabetes agar dapat meningkatkan akurasi prediksi diagnosis diabetes?
2. Bagaimana menemukan model data dengan algoritma *Evolving Artificial Neural Network* untuk memprediksi diagnosis pasien berdasarkan hasil olahan dari set data?
3. Bagaimana analisis *output* dari algoritma *Evolving Artificial Neural Network*?

## 1.3 Tujuan Penelitian

Dari rumusan masalah yang telah dituliskan sebelumnya, terdapat beberapa tujuan dari penelitian ini adalah sebagai berikut:

1. Mengolah dan memproses set data yang digunakan untuk meningkatkan akurasi dari prediksi diagnosis diabetes.

2. Menemukan model data dari algoritma *Evolving Artificial Neural Network* untuk memprediksi diagnosis pasien diabetes.
3. Menganalisis *output* yang dihasilkan dari algoritma *Evolving Artificial Neural Network* dan waktu eksekusi. Juga membandingkan luaran yang dihasilkan dengan menggunakan *software library*.

#### 1.4 Manfaat Penelitian

Adapun manfaat dalam penelitian ini, di antaranya adalah sebagai berikut:

1. Dapat mendeteksi penyakit diabetes dengan hasil dari prediksi sistem, sehingga dapat dilakukan tindakan medis sejak dini.
2. Dapat digunakan sebagai acuan untuk meningkatkan akurasi diagnosis terhadap pasien.
3. Dapat menurunkan angka penderita diabetes dengan melakukan tindakan medis sejak dini sehingga pasien bisa terbebas dari diabetes.

#### 1.5 Batasan Masalah Penelitian

Batasan masalah sangat diperlukan agar pembahasan tidak terlalu luas. Berikut ini merupakan batasan masalah yang terdapat pada penelitian ini.

1. Data yang digunakan merupakan set data terbuka Pima Indian Diabetes. Deskripsi dari setiap kolom yang terdapat di dalam set data ini terdapat pada Tabel 1.1.

Tabel 1.1 Deskripsi kolom pada set data Pima Indian Diabetes

<i>Feature</i>	<b>Deskripsi</b>
<i>Pregnancies</i>	Jumlah berapa kali pernah hamil
<i>Glucose</i>	Konsentrasi plasma glukosa dalam tes toleransi glukosa oral
<i>Blood Pressure</i>	Tekanan darah diastolik (mm Hg)
<i>Skin Thickness</i>	Ketebalan lipatan kulit triceps (mm)

<i>Feature</i>	<b>Deskripsi</b>
<i>Insulin</i>	Serum insulin 2 jam (mu U / ml)
<i>BMI</i>	Indeks massa tubuh (berat dalam kg/(tinggi dalam m) <sup>2</sup> )
<i>Diabetes Pedigree Function</i>	Fungsi silsilah diabetes
<i>Age</i>	Usia (Tahun)
<i>Outcome</i>	Variabel <i>class/boolean</i> 1 dan 0

2. Pengukuran performa pada model akan menggunakan *confusion matrix* yang kemudian akan digunakan untuk menghitung nilai akurasi, *precision*, *recall* dan *f1-score*.

## 1.6 Sistematika Penulisan

Pada bagian sistematika penulisan ini akan diuraikan mengenai penjelasan tiap bab.

### BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang diadakannya penelitian, disusun berdasarkan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

### BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan mengenai teori yang dipelajari selama melaksanakan penelitian. Teori yang terdapat pada bab ini ialah mengenai *machine learning*, *artificial neural network*, algoritma evolusi, prediksi, diabetes, dan penelitian terdahulu.

### BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah penelitian yang akan dilakukan, dimulai dari desain penelitian, fokus penelitian, kemudian alat dan bahan yang digunakan untuk penelitian dan yang terakhir adalah metode penelitian.

### BAB IV HASIL DAN PEMBAHASAN

Bab ini menjabarkan hasil penelitian yang telah dilakukan dan analisisnya. Semua pertanyaan mengenai masalah yang diangkat dalam tema skripsi dibahas di sini. Mulai dari pengumpulan/pengunduhan data, praproses data, pembangunan model *machine learning*, dan analisis hasil serta akurasi dari model.

### BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran bagi peneliti selanjutnya untuk mengembangkan implementasi algoritma pada kasus prediksi diagnosis diabetes dengan menggunakan set data yang digunakan pada penelitian yang telah dilakukan.