

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Autisme atau biasa disebut ASD (*Autistic Spectrum Disorder*) merupakan gangguan perkembangan fungsi otak yang kompleks dan sangat bervariasi (spektrum), biasanya gangguan ini meliputi cara berkomunikasi, berinteraksi sosial dan kemampuan berimajinasi (Pangestu dan Fibriana, 2017).

Autisme dapat didiagnosis pada usia berapa saja, namun gejala umumnya akan terlihat pada anak-anak mulai sejak usia dua tahun. Berdasarkan *National Center on Birth Defects and Developmental Disabilities Centers for Disease Control and Prevention* (2021), Para ilmuwan dari Pusat Pengendalian dan Pencegahan Penyakit (CDC) dan Administrasi Sumber Daya dan Layanan Kesehatan (HRSA) menemukan bahwa 17% anak-anak yang berusia 3-17 tahun memiliki cacat perkembangan dan yang persentase ini meningkat dari tahun ke tahun. Pada dua periode waktu yang dibandingkan, yaitu tahun 2009–2011 dan 2015–2017 terdapat peningkatan yang juga terlihat untuk cacat perkembangan spesifik pada kelompok usia yang sama.

Terdapat beberapa penelitian terdahulu yang menjadi referensi terkait dengan metode yang penulis usulkan, yaitu seperti pada penelitian yang menerapkan *Principal Component Analysis* (PCA) Dalam Penentuan Faktor Dominan Yang Mempengaruhi Pengidap Kanker Serviks dengan menggunakan dataset *Cervical Cancer Dataset* dimana hasil dari penelitian tersebut didapat 9 variabel skrining dengan total varians sebesar 99% (Nasution dkk., 2020). Pada penelitian yang membahas analisis peningkatan akurasi pada *Random Forest* dengan seleksi fitur *Principal Component Analysis* menggunakan dataset *Water Quality Status Dataset*, dapat diketahui nilai hasil akurasi sebesar 91,43% (Dhany, 2020). Kemudian pada penelitian yang melakukan prediksi curah hujan (*Rainfall Dataset*) dengan menggunakan Algoritma *Random Forest* dapat menghasilkan nilai akurasi tertinggi yaitu sebesar 99,45% (Primajaya dan Sari, 2018). Pada penelitian (Surya & Kalpana, 2020) bertujuan untuk menerapkan algoritma *machine learning* KNN terhadap kumpulan data *Autism Spectrum Disorder* yang mana sebelum pembuatan

model, akan dilakukan tahap pembagian data dan praproses data, yaitu ada proses *label encoding* dimana data bertipe kategorikal akan *ditranspose* menjadi *integer*. Pembagian data dilakukan dengan menerapkan rasio perbandingan 7:3, yang berarti diaplikasikan sebesar 70% sebagai data latih dan 30% lainnya untuk data uji, dan didapat hasil penelitian dengan nilai *accuracy* (akurasi) sebesar 90%. Terdapat penelitian lainnya yang menerapkan dataset yang sama yaitu dengan melakukan pemodelan prediktif Gangguan Spektrum Autisme pada Anak-anak menggunakan algoritma C.45 saja. Hasil dari penelitian ini yaitu dapat diketahui nilai akurasi paling tinggi sebesar 92.12% ketika dilakukan pengujian dengan menyertakan fitur A6\_Score (Farhan dan Subekti, 2018). Kemudian pada penelitian yang berbeda namun masih dengan menggunakan dataset yang sama dengan menggunakan algoritma klasifikasi *Naïve Bayes*, diketahui nilai akurasi yang dihasilkan sebesar 60,48% dikarenakan pada praproses data tahap *data cleaning*, penanganan *missing value* dilakukan dengan cara menghapus seluruh *missing value* yang ada pada dataset tanpa memperbaikinya (Rahayu dkk., 2019). Pada penelitian lainnya yang masih berhubungan dengan pendeteksian ASD yaitu ada yang melakukan Perbandingan Akurasi Algoritma C4.5 Dan *Naïve Bayes* Untuk Deteksi Dini Gangguan Autisme Pada Anak, menghasilkan nilai akurasi sebesar 72% yang menggunakan algoritma C4.5 dan menghasilkan sebesar 73.33% yang menggunakan algoritma *Naïve Bayes* (Sugara dkk., 2019).

Berdasarkan beberapa penelitian tersebut, penulis mengusulkan untuk melakukan penelitian yaitu dengan judul Penerapan Klasifikasi *Random Forest* Terhadap Data Gangguan Spektrum Autisme (ASD) pada Anak-anak Menggunakan Metode Seleksi Fitur *Principal Component Analysis* (PCA) dengan Klasifikasi *Random Forest*, dimana metode seleksi fitur PCA akan diterapkan untuk mereduksi dimensi data dan kemudian hasil data dari proses PCA nantinya akan diklasifikasikan pada pemodelan klasifikasi *Random Forest*. Selain itu, akan melihat juga apakah terdapat perbedaan pada performa klasifikasi terhadap data yang dengan menerapkan proses seleksi fitur PCA dengan yang tanpa melalui proses tersebut.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan yang telah diuraikan, dapat dirumuskan permasalahan yang dihadapi sebagai berikut:

- a. Bagaimana menerapkan seleksi fitur menggunakan *Principal Component Analysis* (PCA) dan metode klasifikasi *Random Forest* terhadap Data Gangguan Spektrum Autisme (ASD) pada anak – anak?
- b. Bagaimana hasil evaluasi (*confusion matrix*) dari penerapan metode *Principal Component Analysis* (PCA) dengan menggunakan metode klasifikasi *Random Forest* terhadap Data Gangguan Spektrum Autisme (ASD) pada anak – anak?
- c. Bagaimana perbedaan pada hasil evaluasi (*confusion matrix*) terhadap Data Gangguan Spektrum Autisme (ASD) pada anak – anak antara menggunakan metode seleksi fitur *Principal Component Analysis* (PCA) dengan yang tidak menggunakan metode tersebut?

## 1.3 Ruang Lingkup

Berdasarkan permasalahan yang telah diuraikan memiliki batasan-batasan masalah lain yaitu sebagai berikut:

- a. Dataset yang digunakan pada penelitian ini adalah *Autistic Spectrum Disorder Screening Data for Children Data Set* yang diperoleh dari *University of California Irvine (UCI) Machine Learning Data Repository*.
- b. Dataset yang digunakan mengandung atau terdapat *missing values*.
- c. Dataset memiliki 20 fitur dan 1 atribut target (terdapat 2 label kelas) dengan 292 rekam data.
- d. Metode yang digunakan dalam penelitian ini adalah Klasifikasi *Random Forest* dan dengan menerapkan seleksi fitur *Principal Component Analysis* (PCA).
- e. Bahasa pemrograman yang digunakan dalam penelitian ini adalah menggunakan bahasa pemrograman Python.

- f. Proses pengujian dilakukan dengan dua pendekatan yaitu terhadap data yang melalui proses seleksi fitur *Principal Component Analysis* (PCA) dan tanpa melalui proses tersebut.

#### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan di atas, maka tujuan dari penelitian ini adalah:

- a. Menerapkan metode seleksi fitur *Principal Component Analysis* (PCA) untuk mereduksi dimensi data dari dataset Data Gangguan Spektrum Autisme (ASD) pada Anak-anak serta mengklasifikasikannya menggunakan pemodelan klasifikasi *Random Forest*.
- b. Untuk mengetahui hasil evaluasi (*confusion matrix*) berupa nilai akurasi, *precision*, *recall* dari pemodelan klasifikasi random forest terhadap data Gangguan Spektrum Autisme (ASD) pada Anak – anak yang telah dilakukan proses *Principal Component Analysis* (PCA).
- c. Untuk mengetahui apakah terdapat perbedaan atau tidak pada hasil evaluasi (*confusion matrix*) terhadap Data Gangguan Spektrum Autisme (ASD) pada Anak – anak antara menggunakan metode seleksi fitur *Principal Component Analysis* (PCA) dengan yang tidak menggunakan metode tersebut.

#### 1.5 Manfaat Penelitian

Adapun manfaat yang didapat dari penelitian skripsi ini adalah sebagai berikut:

- a. Mendapatkan pengetahuan bagaimana menerapkan klasifikasi Data Gangguan Spektrum Autisme (ASD) pada Anak-anak dengan menggunakan seleksi fitur *Principal Component Analysis* (PCA) dan metode klasifikasi *Random Forest*.
- b. Mendapatkan informasi mengenai hasil evaluasi (*confusion matrix*) dari penerapan klasifikasi Data Gangguan Spektrum Autisme (ASD) pada anak-anak menggunakan seleksi fitur *Principal Component Analysis* (PCA) dan metode *Random Forest*.

- c. Mendapatkan pengetahuan terhadap ada atau tidaknya perbedaan pada hasil evaluasi (*confusion matrix*) dari klasifikasi *random forest* terhadap Data Gangguan Spektrum Autisme (ASD) pada Anak – anak antara yang menggunakan metode seleksi fitur *Principal Component Analysis* (PCA) dengan yang tidak menggunakan metode tersebut.
- d. Dapat dijadikan sebagai referensi terhadap penelitian selanjutnya dengan topik yang serupa.

## 1.6 Luaran yang Diharapkan

Luaran yang diharapkan dalam penelitian ini adalah berupa performa atau hasil evaluasi dari penerapan metode seleksi fitur *Principal Component Analysis* (PCA) terhadap Data Gangguan Spektrum Autisme pada Anak-anak (*ASD for Children*) menggunakan metode klasifikasi *random forest* dan mengetahui bagaimana pengaruh atau perbedaan dari hasil evaluasi antara data yang menggunakan metode seleksi fitur *Principal Component Analysis* (PCA) dengan yang tidak menggunakan metode tersebut.

## 1.7 Sistematika Penulisan

Sistematika penulisan pada penelitian ini, disusun berdasarkan aturan penulisan yang terdiri dari beberapa bagian, yaitu:

### **BAB I PENDAHULUAN**

Bab ini berisi penjelasan dari latar belakang, rumusan masalah, ruang lingkup, tujuan penelitian, manfaat penelitian, luaran yang diharapkan, dan sistematika penulisan dalam menyusun skripsi.

### **BAB II LANDASAN TEORI**

Pada bab ini berisi uraian mengenai berbagai literatur yang berkenaan dengan teori/konsep/prosedur/metode/proses yang berkaitan dengan penelitian.

### **BAB III METODOLOGI PENELITIAN**

Pada bab ini berisi tahap – tahap penelitian atau metode yang digunakan dalam menyelesaikan permasalahan penelitian sehingga mencapai tujuan penelitian.

### **BAB IV PEMBAHASAN**

Pada bab ini berisi tahap – tahap dari proses yang dilakukan dalam penelitian terhadap permasalahan penelitian sehingga mencapai hasil yang diharapkan dan sesuai dengan tujuan peneliti.

### **BAB V PENUTUP**

Pada bab ini berisi kesimpulan dan saran. Kesimpulan memuat hal-hal yang sudah dibahas pada bab-bab sebelumnya mulai dari permasalahan, analisis sampai dengan hasil penelitian. Saran memuat hal-hal yang perlu dilakukan oleh peneliti selanjutnya (dapat berupa hal-hal yang belum dilakukan oleh penulis dalam penelitian).

### **DAFTAR PUSTAKA**

Pada halaman ini berisi daftar referensi yang digunakan dalam penyusunan proposal. Setiap literatur yang digunakan dalam penyusunan dicantumkan pada halaman ini.

### **RIWAYAT HIDUP**

Pada halaman ini berisi riwayat hidup penulis secara profesional, di dalamnya berisikan tempat, tanggal lahir penulis, nama orang tua penulis, pendidikan penulis, riwayat studi penulis, serta pengalaman organisasi penulis. Dan dilengkapi tanda tangan dan foto penulis.

### **LAMPIRAN**

Pada halaman ini berisi data dan hasil olahan yang mendukung dalam penyusunan Skripsi.