

BAB 1

PENDAHULUAN

1.1.Latar Belakang

Autis adalah kondisi dimana ada masalah kompleks pada gangguan sarafnya. Dampak yang timbul karena adanya masalah pada saraf berupa susah berinteraksi, susah berkomunikasi non-verbal dan berkomunikasi verbal, susah berbicara hingga mengalami kesusahan dalam hal sosial-motorik. Hal ini lah yang dianggap perlu untuk ditangani, berhubung 1 dari 600 anak di Indonesia menyandang penyakit autis, harus ada suatu cara untuk menyelesaikan permasalahan (Budiman, Santoso, & Afirianto, 2017). Ini akan sangat bermanfaat bagi dokter, membantu mereka mendeteksi *Autism Spectrum Disorder* pada tahap yang jauh lebih produktif (Bekerom, 2012). Dikarenakan oleh itu jika seorang dokter dapat menemukan gejala lebih awal maka dapat ditangani lebih awal juga.

Seiring dengan perkembangan teknologi informasi dan ilmu pengetahuan, maka pembelajaran mesin di bidang komputer semakin menarik perhatian. pembelajaran mesin menjadi tren dalam hal teknologi informasi. Penggunaan pembelajaran mesin adalah penggunaan data. *Machine learning* memainkan peran yang luas dalam pengembangan terutama dalam pengembangan data analitik (Alarifi & Young, 2018). Ada beberapa metode dalam pembelajaran mesin dan salah satunya adalah pohon keputusan. Pohon keputusan adalah salah satu dari beberapa metode klasifikasi yang umum digunakan, dan algoritma yang biasa digunakan dalam pohon keputusan adalah C4.5.

Di bidang kesehatan, *machine learning* juga dapat mengklasifikasikan suatu penyakit dari data gejala yang dialami pasien atau dari rekam medis pasien. Algoritma klasifikasi *machine learning* dapat membantu para ahli medis dalam mendiagnosis suatu penyakit atau kelainan pada pasien, contohnya ialah autisme.

Penelitian tentang dataset gangguan autisme pada anak dilakukan oleh (Sugara et al., 2018) dalam penelitiannya menggunakan algoritma C4.5 didapatkan akurasi

sebesar 72%. Akurasi sangat penting dalam klasifikasi dan akurasi 72% masih dapat ditingkatkan.

Metode ensemble adalah metode yang digunakan untuk meningkatkan akurasi algoritma klasifikasi dengan membangun beberapa pengklasifikasi pada data latih dan kemudian pada saat klasifikasi metode ini menggunakan voting/agregasi dari pengklasifikasi tersebut. Salah satu contoh metode ensemble adalah *bootstrap aggregating* yang biasa disingkat dengan *bagging*.

Bootstrap aggregating, disebut juga *bagging*, adalah penggabungan algoritma pembelajaran mesin (*machine learning*) yang dirancang untuk meningkatkan stabilitas dan akurasi dari algoritma *machine learning* yang digunakan dalam klasifikasi statistik dan regresi. *Bagging* juga mengurangi varians dan membantu untuk menghindari terjadinya *overfitting*. Meskipun biasanya diterapkan untuk metode decision tree, *bagging* dapat digunakan dengan semua jenis metode. *Bagging* merupakan kasus khusus dari pendekatan model *averaging*. *Bagging* dapat digunakan untuk klasifikasi dan regresi.

Metode *bagging* membentuk kelas algoritma yang membangun beberapa contoh *estimator* kotak hitam pada himpunan bagian acak dari training set asli dan kemudian menggabungkan prediksi masing-masing untuk membentuk prediksi akhir.

Bagging merupakan salah satu metode yang berdasar pada ensemble method, oleh karena itu secara umum tahap-tahap pada metode *bagging* memiliki dua hal yang harus dilakukan yaitu proses *bootstrap* dan *aggregating*. Ada beberapa hal penting dalam metode ini yaitu :

- a) Pendistribusian data (*bootstrap*) dibuat dengan menggunakan *sampling with replacement*.
- b) Membangun classifier pada setiap *bootstrap sample*.
- c) Mendapatkan model pada *classifier* dan melakukan *aggregating (voting)*.

Pada penelitian yang dilakukan oleh (Sulaiman & Laksana, 2015) menggunakan algoritma C4.5 dan C4.5 berbasis *bagging* yang akan digunakan untuk menganalisis kepuasan pelanggan, dimana data retensi menggunakan 460 pelanggan pulsa

elektronik dan menggunakan 11 atribut. Dalam penelitiannya, algoritma information gain C4.5 memperoleh akurasi sebesar 84,8% dan ketika bagging digunakan pada algoritma information gain C4.5 akurasinya meningkat 6% menjadi 90,8%. Dalam penjelasannya, penelitian yang akan dilakukan bertujuan untuk meningkatkan akurasi algoritma C4.5 pada data gangguan autisme dengan menggunakan teknik bagging.

1.2.Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan diatas, rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Berapa tingkat presisi algoritma C4.5 sebelum dan sesudah menggunakan teknik *bagging* pada data gangguan autisme.
2. Berapa tingkat akurasi algoritma C4.5 sebelum dan sesudah menggunakan teknik *bagging* pada data gangguan autisme.

1.3.Batasan Penelitian

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunnakan merupakan data dari penelitian yang dilakukan oleh (Sugara et al., 2018).
2. Data yang dihitung sebagai *sample* adalah data gejala-gejala autisme dengan jumlah record sebanyak 50 data.
3. Penentuan gangguan autisme berdasarkan 24 parameter yaitu :
 - GJ01 : Tidak memiliki kontak mata
 - GJ02 : Suka diam/menyendiri
 - GJ03 : Tidak suka dipeluk
 - GJ04 : Tidak dapat merespon jika dipanggil orang
 - GJ05 : Suka melakukan kegiatan/gerakan secara berulang-ulang
 - GJ06 : Suka terpaku terhadap benda-benda tertentu
 - GJ07 : Suka menyukai hal yang aneh seperti mencium-cium benda
 - GJ08 : Suka mengungkapkan emosi (sedih, senang, marah dll) dengan sendirinya

tanpa sebab

GJ09 : Tidak bisa diam

GJ10 : Tidak dapat berbicara

GJ11 : Bisa berbicara namun tidak jelas

GJ12 : Sering berbicara berlebihan

GJ13 : Suka mengucapkan bahasa/kata-kata yang aneh secara berulang-ulang

GJ14 : Tidak dapat menunjuk sesuatu dengan jari sendiri

GJ15 : Tidak dapat menunjukkan keinginan dengan kata-kata

GJ16 : Suka menarik-narik orang lain jika menginginkan sesuatu

GJ17 : Tidak ada usaha dalam berkomunikasi

GJ18 : Menghindar jika didekati

GJ19 : Tidak dapat berinteraksi dengan lingkungan sekitar

GJ20 : Tidak tertarik dengan orang lain

GJ21 : Tidak peduli dengan sekitarnya

GJ22 : Tidak suka dengan keramaian

GJ23 : Tidak suka bermain dengan teman sebayanya

GJ24 : Tidak dapat bersosialisasi dengan orang lain

4. Output yang digunakan pada penelitian ini ada 3 yaitu Gangguan Perilaku, Gangguan Komunikasi, Gangguan Interaksi Sosial.
5. *Tools* yang akan digunakan adalah “*RapidMiner Studio*”.

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Mengetahui presisi algoritma C4.5 sebelum dan sesudah menggunakan teknik *bagging* pada data gangguan autisme.
2. Mengetahui akurasi algoritma C4.5 sebelum dan sesudah menggunakan teknik *bagging* pada data gangguan autisme.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Dapat digunakan untuk alternatif penggunaan klasifikasi C4.5.
2. Sebagai peningkatan presisi dan akurasi algoritma C4.5.

