

1. Pendahuluan

Autistic Spectrum Disorder atau disingkat ASD adalah kondisi perkembangan syaraf yang tidak sempurna yang terjadi pada banyak remaja di dunia[1]. Kasus ini terjadi pada negara maju dan negara berkembang. Di Indonesia sudah terdapat kasus remaja yang mengidap penyakit ASD. Salah satu cara untuk mengobservasi penyakit ini dengan mengadakan serangkaian tes *screening* terhadap remaja yang bersangkutan. Metode atau tes yang digunakan tersebut salah satunya adalah mengajukan pertanyaan-pertanyaan yang diberikan oleh mereka yang kompeten dengan penyakit ini. Jawaban dan identitas remaja tersebut akan menjadi suatu data untuk mengklasifikasi apakah remaja tersebut mengidap penyakit ASD atau tidak.

Masalah utama yang dihadapi adalah metode *screening* yang membutuhkan biaya yang besar. Biaya yang besar disebabkan oleh observasi pada penderita yang lama dilakukan karena berhubungan dengan karakter dan perilaku dari setiap remaja yang menjadi subjek penelitian. Karakter dan perilaku adalah suatu hal yang harus diamati berulang-ulang sehingga membutuhkan biaya besar dalam melakukan observasi. Dampak ekonomi dari autisme dan peningkatan jumlah kasus ASD di seluruh dunia menunjukkan adanya kepentingan yang mendesak untuk mengembangkan metode *screening* yang lebih efektif. Bagi para peneliti, data yang terkait dalam kasus ASD tidak banyak. Pada beberapa penelitian ASD, metode *machine learning* menggunakan *k-nearest neighbor* (kNN) dan *The Linear Discriminant Analysis* (LDA), Metode kNN menghasilkan akurasi sebesar 88,5%[2] dan metode LDA menghasilkan akurasi sebesar 90,8%[3].

Latar Belakang

Pada jurnal – jurnal yang menyangkut penelitian tentang kasus kasus ASD, metode pengujian *machine learning* menggunakan menggunakan *random forest* menghasilkan akurasi 100%[4]. Kondisi akurasi 100% didapatkan ketika data *training* sama dengan data *testing*, karena semua data sudah berada dalam pohon keputusan sehingga *random forest* dapat mengambil keputusan yang tepat.

Pada data ASD yang digunakan, fitur – fitur pada data tersebut bernilai diskrit. Atribut diskrit cocok digunakan untuk metode *random forest* karena metode *random forest* dapat mengklasifikasikan data diskrit dengan baik[9]. Akan tetapi akurasi dalam *random forest* pada beberapa penelitian bersifat *overfit*. *Random forest* bersifat *overfit*[5] dikarenakan apabila data *testing* tidak ada dalam suatu pohon keputusan voting, maka sistem *random forest* akan sulit dalam mengklasifikasikan data.

Untuk menangani permasalahan *random forest* yang bersifat *overfit*, digunakanlah algoritma optimasi *firefly algorithm*. Pada beberapa kasus optimasi data diskrit dengan algoritma *firefly algorithm*, algoritma tersebut dapat membuat akurasi menjadi lebih optimal yang ditandai dengan nilai akurasi yang semakin tinggi[10]. Dengan menggunakan algoritma optimasi *firefly algorithm*, dapat membuat metode *random forest* menjadi tidak *overfit* dan bersifat normal.

Optimasi yang dilakukan dan metode klasifikasi akan menentukan kelas pada data ASD yang baru saat dilakukan observasi.

Topik dan Batasannya

Firefly Algorithm akan diterapkan pada *random forest* untuk optimasi akurasi dan *F1-score*. Representasi individu pada *firefly* adalah satu individu adalah satu pohon keputusan. Representasi optimasi adalah pergerakan *firefly* berdasarkan intensitas cahaya yang ada. Intensitas cahaya berdasarkan akurasi setiap *firefly*, semakin besar nilai akurasi pada suatu *firefly*, maka semakin besar pula nilai intensitas cahaya.

Pada tugas akhir yang dikerjakan, terdapat beberapa batasan yang digunakan. Data yang digunakan terbatas pada data *autistic spectrum disorder* yang diterbitkan oleh UCI *machine learning* dan berasal dari penelitian yang dilakukan oleh Manukau Institute of Technology[8]. Untuk memvalidasi digunakan metode *k-fold* dan metode pengambilan sampel dengan menggunakan *bootstrap*. Atribut data pada ASD bersifat diskrit. Metode klasifikasi yang digunakan adalah *random forest*. Algoritma optimasi yang digunakan adalah *firefly algorithm*.

Tujuan

Tujuan akhir dari digunakannya *firefly algorithm* pada metode *random forest* adalah terbentuknya pohon keputusan yang telah dioptimasi dan pohon keputusan tersebut akan berpengaruh kepada nilai akurasi yang semakin besar. Nilai akurasi yang besar akan berpengaruh terhadap *random forest*. Sistem klasifikasi *random forest* yang tidak *overfit* akan terbentuk untuk mengklasifikasikan data ASD secara optimal dan sistem ini berguna sebagai landasan para peneliti.