



ABSTRACT

Early detection of ASD (Autism Spectrum Disorder) can help people know the properties or symptoms associated with ASD early. Early detection of ASD has been done by researchers with classification techniques using machine learning. But the results of classifications using machine learning sometimes do not necessarily provide good performance. Poor performance is often caused by all features that do not contribute positively to the classification. Therefore, it is necessary to do a feature selection to get a subset of features that can improve classification performance. The purpose of this study is to improve the performance of ASD classification using wrapper-based feature selection methods. The wrapper-based feature selection method applies a method of finding a subset of features, namely the Sequential Forward Selection algorithm and classifier such as SVM, Random Forest (RF), and ANN. Whereas the classification process uses the JRip and C4.5 classifier. Using the wrapper-based feature selection method will get a subset of features that contribute positively to improve ASD classification performance.

The *dataset* used in this study is AQ-10-Child and AQ-10-Adult. In this study comparing the modelling performance in 2 schemes: first, the comparison of performance between classification methods without feature selection with the proposed method. Second, the comparison of performance between previous research methods with the proposed method. The experimental results show that in the first scheme there are 2 proposed methods, namely: SFSSVM + JRip and SFSRF + JRip with 8 subsets of features found to have higher performance compared to the 10 feature classification results of the JRip algorithm on the AQ-10-Child *dataset*. The performance results of the SFSSVM + JRip and SFSRF + JRip methods both obtained an accuracy performance of 90.44%, sensitivity 87.39%, and specificity 93.72%. Whereas in the second scheme, the three methods proposed with 8 subsets of features obtained had a positive effect on improving the performance of the accuracy, sensitivity, and specificity of previous



studies in the AQ-10-Child and AQ-10-Adult *datasets*. The highest performance results obtained by the SFSSVM + C4.5 and SFSRF + C4.5 methods both produce an accuracy of 90.64%, a sensitivity of 90.85% and a specificity of 90.42% in the AQ-10-Child *dataset*. Whereas in the AQ-10-Adult *dataset* the highest performance was obtained by the SFSRF + JRip method with an accuracy of 93.07%, sensitivity 94.08%, and specificity 90.30% and SFSRF + C4.5 with an accuracy of 91.85%, sensitivity 94 , 77%, and specificity 83.87%. Statistical test results show that the two proposed methods (SFSSVM + JRip and SFSRF + JRip) have significant differences in accuracy and sensitivity performance of the classification method without feature selection on the AQ-10-Child *dataset*. Whereas in the second scheme, the three proposed methods get a significant difference in performance from the previous research methods in the two *datasets*.

Keywords : ASD, *Feature selection, wrapper, klasifikasi ASD, Sequential Forward Selection, SVM, RF, ANN, JRip, C4.5.*



INTISARI

Pendeteksian dini ASD (*Autism Spectrum Disorder*) dapat membantu masyarakat mengetahui sifat-sifat atau gejala-gejala yang berkaitan dengan ASD lebih awal. Pendeteksian dini ASD sudah banyak dilakukan para peneliti dengan teknik klasifikasi menggunakan *machine learning*. Namun hasil dari klasifikasi menggunakan *machine learning* terkadang belum tentu memberikan kinerja yang baik. Kinerja yang kurang baik sering kali disebabkan oleh semua fitur yang tidak berkontribusi positif dalam klasifikasi. Oleh karena itu, perlu dilakukan seleksi fitur untuk mendapatkan *subset* fitur yang dapat meningkatkan kinerja klasifikasi. Tujuan penelitian ini adalah untuk meningkatkan kinerja klasifikasi ASD menggunakan metode seleksi fitur berbasis *wrapper*. Metode seleksi fitur berbasis *wrapper* menerapkan metode *pencarian* subset fitur yaitu algoritme *Sequential Forward Selection* dan *classifier* seperti: SVM, *Random Forest* (RF), dan ANN. Sedangkan pada proses klasifikasi menggunakan *classifier* JRip dan C4.5. Dengan penggunaan metode seleksi fitur berbasis *wrapper* akan mendapatkan subset fitur yang berkontribusi positif dalam meningkatkan kinerja klasifikasi ASD.

Dataset yang digunakan pada penelitian ini yaitu: *AQ-10-Child* dan *AQ-10-Adult*. Pada penelitian ini membandingkan kinerja pemodelan dalam 2 skema: pertama, perbandingan kinerja antara metode klasifikasi tanpa seleksi fitur dengan metode yang diusulkan. Kedua, perbandingan kinerja antara metode penelitian sebelumnya dengan metode yang diusulkan. Hasil eksperimen menunjukkan bahwa pada skema pertama terdapat 2 metode yang diusulkan yaitu: SFSSVM+JRip dan SFSRF+JRip dengan 8 subset fitur yang didapatkan memiliki kinerja lebih tinggi dibandingkan dengan hasil klasifikasi 10 fitur dari algoritme JRip pada *dataset AQ-10-Child*. Hasil kinerja dari metode SFSSVM+JRip dan SFSRF+JRip sama-sama memperoleh kinerja akurasi sebesar 90,44%, sensitivitas 87,39%, dan spesifisitas 93,72%. Sedangkan pada skema kedua, ketiga metode yang diusulkan dengan 8 subset fitur yang didapatkan berpengaruh positif terhadap peningkatan kinerja akurasi, sensitivitas, dan spesifisitas penelitian



sebelumnya pada *dataset AQ-10-Child* dan *AQ-10-Adult*. Hasil kinerja tertinggi diperoleh oleh metode SFSSVM+C4.5 dan SFSRF+C4.5 yang sama-sama menghasilkan akurasi 90,64%, sensitivitas 90,85% dan spesifisitas 90,42% pada *dataset AQ-10-Child*. Sedangkan pada *dataset AQ-10-Adult*, kinerja tertinggi diperoleh oleh metode SFSRF+JRip dengan akurasi 93,07%, sensitivitas 94,08%, dan spesifisitas 90,30% dan SFSRF+C4.5 dengan akurasi sebesar 91,85%, sensitivitas 94,77%, dan spesifisitas 83,87%. Hasil pengujian statistik menunjukkan bahwa 2 metode yang diusulkan (SFSSVM+JRip dan SFSRF+JRip) memiliki perbedaan kinerja akurasi dan sensitifitas yang signifikan dari metode klasifikasi tanpa seleksi fitur pada *dataset AQ-10-Child*. Sedangkan pada skema kedua, ketiga metode yang diusulkan mendapatkan perbedaan kinerja yang signifikan dari metode penelitian sebelumnya pada kedua *dataset*.

Kata kunci -- ASD, seleksi fitur, *wrapper*, klasifikasi ASD, *Sequential Forward Selection*, SVM, RF, ANN, Jrip, C4.5