

Autism Spectrum Disorder Screening Using Machine Learning

Mahit Tanikella

Stratford Middle School – Sunnyvale

Abstract

Objectives/Goals

Autistic Spectrum Disorder (ASD) is a neurodevelopment disorder characterized by impaired communication, cognitive, and social skills and abilities. Existing screening tools for detection of autism are expensive, cumbersome, time-intensive, and sometimes fall short in predictive value. Study of applying different supervised, classification-based machine learning algorithms to the problem of detecting ASD based on simple behavioral traits has not been done before. The objective of this project is to build a machine learning algorithm that can predict with close to 100% accuracy, whether a person has ASD, based on behavioral traits. This low cost, quick and easy to use diagnostic test will help healthcare professionals and individuals assess if they should pursue formal treatment options.

Methods/Materials

Machine learning models were developed for five different algorithms namely Logistic Regression (LR), Decision Trees (DT), Gaussian Naive Bayes (NB), Support Vector Machines (SVM) and Neural Networks (NN). Coding was done in Python using scikit-learn in Jupyter notebooks. The models were trained using ASD screening data from UC Irvine machine learning repository. Data consists of response to questions on behavioral traits, age, gender, ethnicity, if the person had jaundice when born and if anyone in the family is diagnosed with ASD. Data preprocessing was done by cleaning and removing irrelevant fields. Then, data was transformed using label encoding technique which converts non-numerical labels to numerical labels and normalizes the data. For evaluating the models, 10-fold cross validation technique was used in which the data is partitioned into ten equal sizes and nine samples were used for training and one for validation, repeating the process ten times with different sample for validation each time. Each model's accuracy score, confusion matrix that describes performance of the model and classification report were generated.

Results

The models developed have achieved average accuracies of 96.7%(LR), 90.7%(DT), 95.4%(NB), 92.3%(SVM) and 97.4%(NN) with standard deviations of 0.023(LR), 0.036(DT), 0.034(NB), 0.035(SVM) and 0.01(NN) respectively. Neural Networks based model is the best with highest possible accuracy and lowest variance. Dropping gender and age from the input feature list improved accuracy which means they are not useful for predicting ASD. Accuracy of models drop if only response to questions i.e. behavioral traits are used for training. So, other factors i.e. if the person was born with jaundice, family history and ethnicity are important factors to consider along with behavioral traits for ASD screening.

Conclusions/Discussion

Machine learning model developed predicts if someone has ASD with 97.4% accuracy, based on answers to behavioral traits questions. This model can be used as initial screening test before doing more expensive diagnostic tests. People can take this test from the comfort of their home, on their computer or mobile phone for initial assessment. By collecting more data and using that for training, we can further improve the accuracy of the model.