Name(s) Mahit V. Tanikella	Project Number **J1508**

Project Title

Autism Spectrum Disorder Screening Using Machine Learning

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 and time-intensive. The objective of this project is to create a low cost, quick and easy to use diagnostic test for ASD by building a machine learning algorithm that can predict with close to 100% accuracy, whether a person has ASD, based on behavioral traits.

Methods/Materials

Machine learning models were developed for five different classification 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 notebook. 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, family history and if the person had jaundice when born. 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. Accuracy score, confusion matrix that describes performance of the model and classification report were generated for each model.

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. Family history, ethnicity and if the person was born with jaundice are important factors to consider along with behavioral traits for ASD screening.

Conclusions/Discussion

Neural networks based machine learning model developed predicts if someone has ASD with 97.4% accuracy, based on answers to behavioral traits questions. People can take this test from the comfort of their home, on their computer or mobile phone for initial assessment, before doing more expensive diagnostic tests.

Summary Statement

Neural networks based machine learning model developed predicts if someone has Autism Spectrum Disorder with 97.4% accuracy, based on answers to behavioral traits questions.

Help Received

My mom helped me with questions I had when I did the online course to learn how to use scikit-learn machine learning library.