Juan E. Garcia  
Texas A&M-San Antonio  
College of Business   
CISA 4358 Senior Project and Seminar   
Fall 2022   
Project Progress Report #1

1. Introduction that includes the problem definition and description as well as the Motivation (why solving this problem is important).

The dataset will be analyzed to improve the prediction of the likelihood of adults having autism disorders. The dataset is a classification problem. The testing and training dataset provides predictive analysis of an adult who has Autism. By using predictive modeling, we can determine if adults have Autism. There are a few critical descriptions that will potentially give better predictions. For example, gender will be a significant column in ensuring that it is clean and essential to the dataset. While Autism affects all genders, current worldwide data suggest that it is more prominent in boys or men. The collected data must remove the gender bias and determine the diagnoses based on other factors(site). Another column of significant factors will be the Autism column, which signifies whether there is a direct relative on the spectrum. According to the CDC, ASD (autism spectrum disorder), if you have family members on the spectrum, it increases the chances of you having Autism. Lastly, the score from which Autism Spectrum Quotient (AQ) is separated into ten questions and is a guideline that covers diagnosing and managing suspected or confirmed autism spectrum disorder in people aged 18 and over. Depending on the score, it would assist in the prediction of Autism in adults.

The motivation to solve this prediction model comes from the fact that I have a daughter diagnosed with Autism. Luckily for us, we caught it when she was three years old. However, my brother-in-law did not know he was autistic until after several years of medication that perhaps could have been avoided and several behavioral psychologists that were not fully aware of what was happening in his head. Furthermore, Autism impairs how people learn and communicate and affects their social skills. We can move testing and training models on younger people by using machine learning models to help catch it early. Early diagnoses are why catching Autism as early as possible can get individuals the help they need to be productive citizens worldwide.

Resources

Butter, E. (2019, June 28). *Gender identity and autism*. Autism Speaks. Retrieved October 2022, from <https://www.autismspeaks.org/expert-opinion/gender-identity-and-autism>

Buchanan, D. B. (2022, May 19). *Autism spectrum quotient (AQ)*. NovoPsych. Retrieved October 2022, from https://novopsych.com.au/assessments/diagnosis/autism-spectrum-quotient/

Centers for Disease Control and Prevention. (2022, May 18). *Autism spectrum disorder, family health history, and Genetics*. Centers for Disease Control and Prevention. Retrieved October 19, 2022, from https://www.cdc.gov/genomics/disease/autism.htm

*Overview: Autism spectrum disorder in adults: Diagnosis and management: Guidance*. NICE. (2021, June). Retrieved October 2022, from https://www.nice.org.uk/guidance/CG142

2. Description of the data set you are going to use (the size of data and description of the variables/features contained in the data). Please provide the link of that dataset.

The total size of the dataset is 100.48kb it contains training data set and a testing dataset in a .CSV

The Columns are labeled with a predictor or target column and they are labeled as follows.

* ID - ID of the patient
* A1\_Score to A10\_Score - Score based on Autism Spectrum Quotient (AQ) 10 item screening tool
* age - Age of the patient in years
* gender - Gender of the patient
* ethnicity - Ethnicity of the patient
* jaundice - Whether the patient had jaundice at the time of birth
* autism - Whether an immediate family member has been diagnosed with autism
* contry\_of\_res - Country of residence of the patient
* used\_app\_before - Whether the patient has undergone a screening test before
* result - Score for AQ1-10 screening test
* age\_desc - Age of the patient
* relation - Relation of patient who completed the test
* Class/ASD - Classified result as 0 or 1. Here 0 represents No and 1 represents Yes. This is the target column, and during submission submit the values as 0 or 1 only.

The link to the dataset can be reached at the following URL. <https://www.kaggle.com/competitions/autismdiagnosis/data>