**Data Set Name: Autistic Spectrum Disorder Screening Data for Adult – Version 2**

**Date: FEB, 15, 2018.**

**Author: Dr Fadi Thabtah**

**Abstract:** Autistic Spectrum Disorder (ASD) is a neurodevelopment condition associated with significant healthcare costs, and early diagnosis can significantly reduce these. Unfortunately, waiting times for an ASD diagnosis are lengthy and procedures are not cost effective. The economic impact of autism and the increase in the number of ASD cases across the world reveals an urgent need for the development of easily implemented and effective screening methods. Therefore, a time-efficient and accessible ASD screening is imminent to help health professionals and inform individuals whether they should pursue formal clinical diagnosis. The rapid growth in the number of ASD cases worldwide necessitates datasets related to behaviour traits. However, such datasets are rare making it difficult to perform thorough analyses to improve the efficiency, sensitivity, specificity and predictive accuracy of the ASD screening process. Presently, very limited autism datasets associated with clinical or screening are available and most of them are genetic in nature. Hence, we propose a new dataset related to autism screening of adults that contained over 20 features to be utilised for further analysis especially in determining influential autistic traits and improving the classification of ASD cases. In this dataset, we record ten behavioural features (AQ-10-Adult) plus ten individuals characteristics that have proved to be effective in detecting the ASD cases from controls in behaviour science.

**Source:** Fayez Thabtah

Department of Digital Technology

Manukau Institute of Technology,

Auckland, New Zealand

[fadi.fayez](mailto:fadi.fayez)@manukau.ac.nz

**Data Type: Predictive and Descriptive:** Nominal / categorical, binary and continuous

**Task:** Classification

**Attribute Type:** Categorical, continuous and binary

**Area:** Medical, health and social science

**Format Type:** Non-Matrix

**Does your data set contain missing values?** Yes

**Number of Instances (records in your data set):** 1118

**Number of Attributes (fields within each record):** 23

**Attribute Information:** For Further information about the attributes/feature see below table.

**Attributes:**

A1-A10: Items within [AQ-10 Autistic screening methods](file:///C:\Users\st2730\Desktop\Research\Autism\MS HUDD\Conference paper 2\slightly agree)  in which “slightly agree” & “definitly agree” items’ values are mapped to “1” wheras “slightly disagree” & “definitly disagree” values are mapped to “0” in the datasets. The remaining features in the datasets are collected from the “submit” screen in the [ASDTests screening app](https://play.google.com/store/apps/details?id=com.asd.asdquiz&hl=en). It should be noted that the class varaible was assigned automatically based on the score obtained by the user while undergoing the screening process using the ASDTests app.

**Relevant Papers:**

1) Tabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp.1-6. Taichung City, Taiwan, ACM.

2) Thabtah, F. (2017). ASDTests. A mobile app for ASD screening. www.asdtests.com [accessed December 20th, 2017].

3) Thabtah, F. (2017). Machine Learning in Autistic Spectrum Disorder Behavioural Research: A Review. Informatics for Health and Social Care Journal.

4) Thabtah F. (2018) An Accessible and Efficient Autism Screening Method for Behavioural Data and Predictive Analyses. To Appear in Health Informatics Journal. 2018.

Table 1: Details of variables mapping to the screening methods

|  |  |
| --- | --- |
| **Variable in Dataset** | **Corresponding AQ-10-Adult Features** |
| A1 | I often notice small sounds when others do not |
| A2 | I usually concentrate more on the whole picture rather than the small details |
| A3 | I find it easy to do more than one thing at  once |
| A4 | If there is an interruption, I can switch back to what I was doing very quickly |
| A5 | I find it easy to ‘read between the lines’ when someone is talking to me |
| A6 | I know how to tell if someone listening to me is getting bored |
| A7 | When I’m reading a story I find it difficult to work out the characters’ intentions |
| A8 | I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc) |
| A9 | I find it easy to work out what someone is thinking or feeling just by looking at their face |
| A10 | I find it difficult to work out people’s intentions |

Table 2: Features collected and their descriptions

|  |  |  |
| --- | --- | --- |
| **Feature** | **Type** | **Description** |
| Age | Number | Toddlers (months), children, adolescent, and adults( year) |
| Gender | String | Male or Female |
| Ethnicity | String | List of common ethnicities in text format |
| Born with jaundice | Boolean (yes or no) | Whether the case was born with jaundice |
| Family member with PDD | Boolean (yes or no) | Whether any immediate family member has a PDD |
| Who is completing the test | String | Parent, self, caregiver, medical staff, clinician ,etc. |
| Country of residence | String | List of countries in text format |
| Used the screening app before | Boolean (yes or no) | Whether the user has used a screening app |
| Screening Method Type | Integer (0,1,2,3) | The type of screening methods chosen based on age category (0=toddler, 1=child, 2= adolescent, 3= adult) |
| Language | String | (English, Arabic, Farsi, Mandarin, Urdu, Swahili, French, Spanish, Portuguese, Turkish) |
| Why\_are\_you\_taken\_the\_screening | String | Use input textbox |
| Question 1 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 2 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 3 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 4 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 5 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 6 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 7 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 8 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 9 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Question 10 Answer | Binary (0, 1) | The answer code of the question based on the screening method used |
| Screening Score | Integer | The final score obtained based on the scoring algorithm of the screening method used. This was computed in an automated manner |
| Class | String | ASD traits or No ASD traits (automatically assigned by the ASDTests app). |