Autistic Spectrum Disorder (ASD) cases

Description and Preprocessing of Data

The Autism-Adult-Data.arff is made up of 704 rows and 21 columns. The columns are as follows:

- A1_Score - Question 1 Answer: Binary (0, 1)
- A2_Score - Question 2 Answer: Binary (0, 1)
- A3_Score - Question 3 Answer: Binary (0, 1)
- A4_Score - Question 4 Answer: Binary (0, 1)
- A5_Score - Question 5 Answer: Binary (0, 1)
- A6_Score - Question 6 Answer: Binary (0, 1)
- A7_Score - Question 7 Answer: Binary (0, 1)
- A8_Score - Question 8 Answer: Binary (0, 1)
- A9_Score - Question 9 Answer: Binary (0, 1)
- A10_Score - Question 10 Answer: Binary (0, 1)
- Age - Age in years
- Gender - Gender (m: Male, f: Female)
- Ethnicity - List of common ethnicities   (White-European, Latino, Others, Black, Asian, Middle Eastern, Pasifika,   South Asian, Hispanic, Turkish). I removed this variable because there is not enough data to recognise patterns between ethnicities
- Jundice - Whether the case was born with   Jundice (Yes, No)
- Austim - Whether any immediate family   member has a PDD (Yes, No)
- Country_of_res - Country of residence (List of   countries)
- Used_app_before - Whether the user has used the   screening app before (Yes, No). A high percentage of this variable was no, therefore this variable was dropped for lack of predictive power.
- Result - Screening score: The final score   obtained based on the scoring algorithm of the screening method used. This   was computed in an automated manner
- Age_desc - Age description. This variable was dropped as a high number was extremely similar.
- Relation - Who is completing the test (Self, Parent, Health care professional, Relative, etc). A high percentage of the relations were self, so this variable would not have great predicting power.
- Class/ASD - yes, no

Project Description

The aim for this project was to predict whether an adult has Autism using a Machine Learning Classification Model. The Model created was able to correctly predict approximately 98% of data it had not seen.