Analysis and Detection of Autism Spectrum Disorder for Children

Using Supervised and Unsupervised Machine Learning Techniques

Anicoche - Pacis - Tuazon - Villaluz

Introduction

01 Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects social functioning, communication and the behavior of an individual.

02 Can be diagnosed at any stage of a person's life and is considered to be a behavioral disease/disorder

03 A need for fast and effective screening procedures that are easily accessible to assist health professionals and inform individuals

Objectives

WHAT WE WANT TO ACHIEVE

To be able to use a new type of dataset that records ten behavioral features (AQ-10-Child) plus ten individual characteristics that are relevant and considered as important factors in detecting ASD from controls in behaviour science.

Use of the supervised and unsupervised machine learning models in order to improve the accuracy and detection of ASD in patients.

Dataset Description

01 Number of Instances: 292

02 Number of Attributes: 21

03 Data Type: Categorical, continuous and binary

Dataset Description

Attributes	Type	Description
Age	Number	Age of the patient in years.
Gender	String	Male or Female
Ethnicity	String	A list of common ethnicities written in text format.
Born with Jaundice	Boolean (yes or no)	Whether the patient was born with jaundice at birth.
Family member with PDD	Boolean (yes or no)	Whether the patient has an immediate family member with PDD.
Who is completing the test	String	Whether self, parent, caregiver, medical staff, clinician, etc. completed the test.
Country of residence	String	A list of countries in text format.
Used the screening app before	Boolean (yes or no)	Whether the user has used the screening app before.
Screening Method Type	Integer (0,1,2,3)	Type of screening methods chosen based on age category (0=toddler, 1=child, 2= adolescent, 3= adult)
Question 1 - Question 10 Answers	Binary (0,1)	The answer of the user to the screening questions.
Screening Score	Integer	The final score obtained based on the scoring algorithm of the screening method used that was computed in an automated manner.

Preprocessing Techniques

- The dataset is transformed into a proper and understandable format for our machine learning model.
- Noise, missing values, or errors are typically found when dealing with real-world data. As with our dataset, noise and missing values were present.
- The preprocessing technique performed gets the mean of the column index of the missing values, which is the age attribute.
- 12 columns were dropped from the base data, i.e. A1 to A10 questions, “age_desc” , and "used_app_before", for we thought that they are not relevant to the models

Exploratory Data Analysis

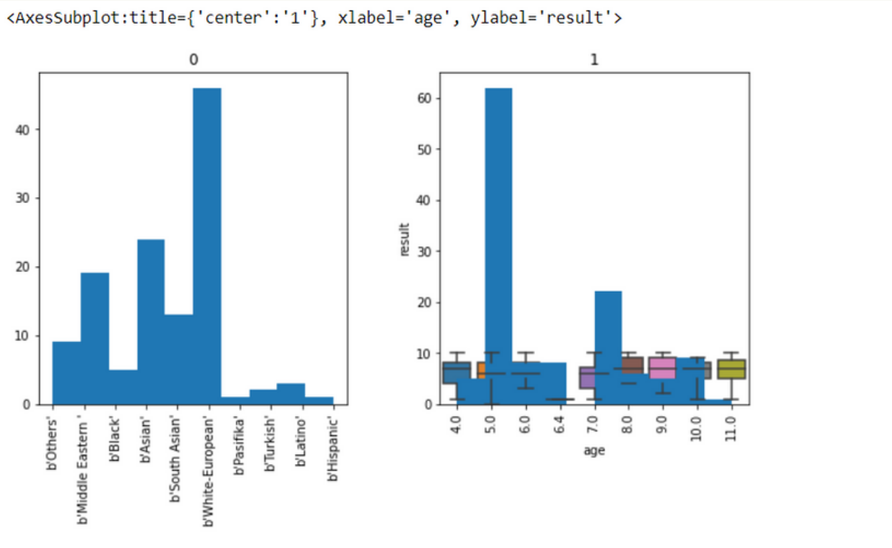


TABLE 1. ETHNICITY AND AGE

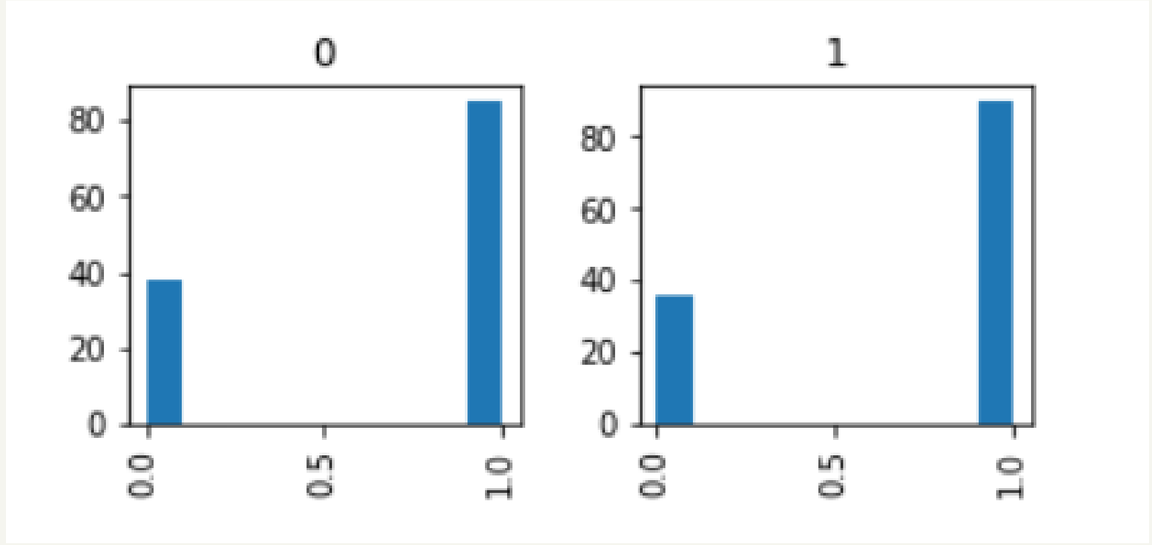


TABLE 2. GENDER

Exploratory Data Analysis

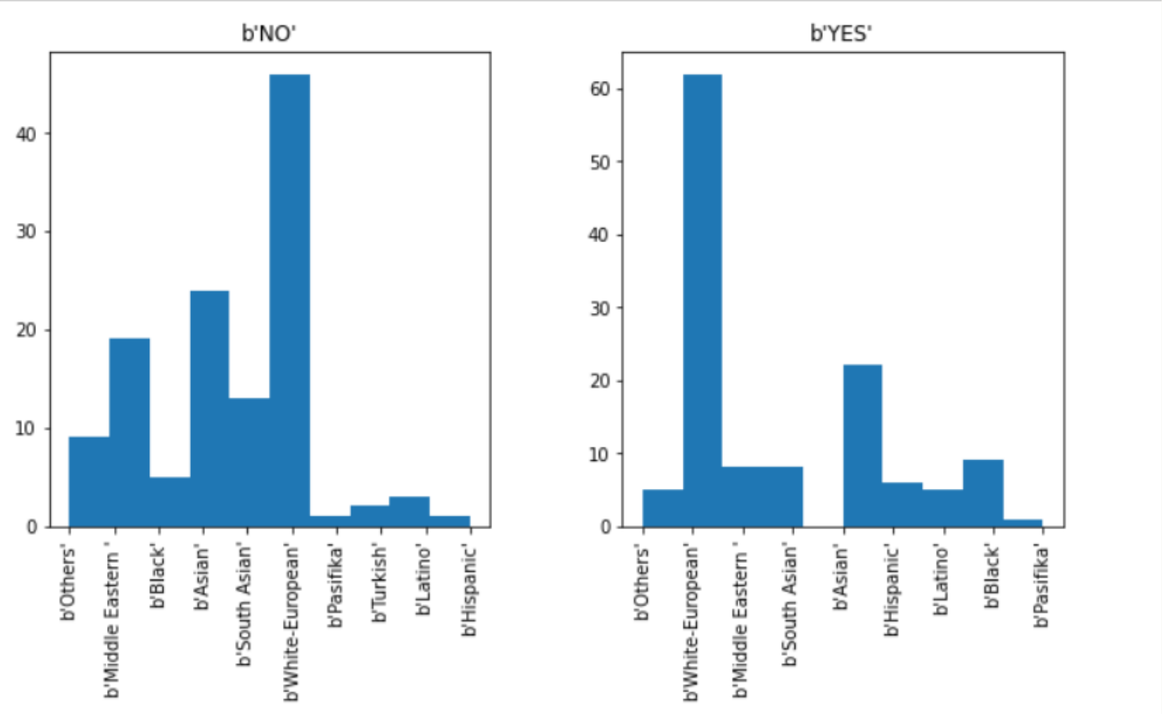


TABLE 3. ETHNICITIES WITH/WITHOUT ASD

Machine Learning Models

CMSC 176: FUNDAMENTALS OF DATA SCIENCE

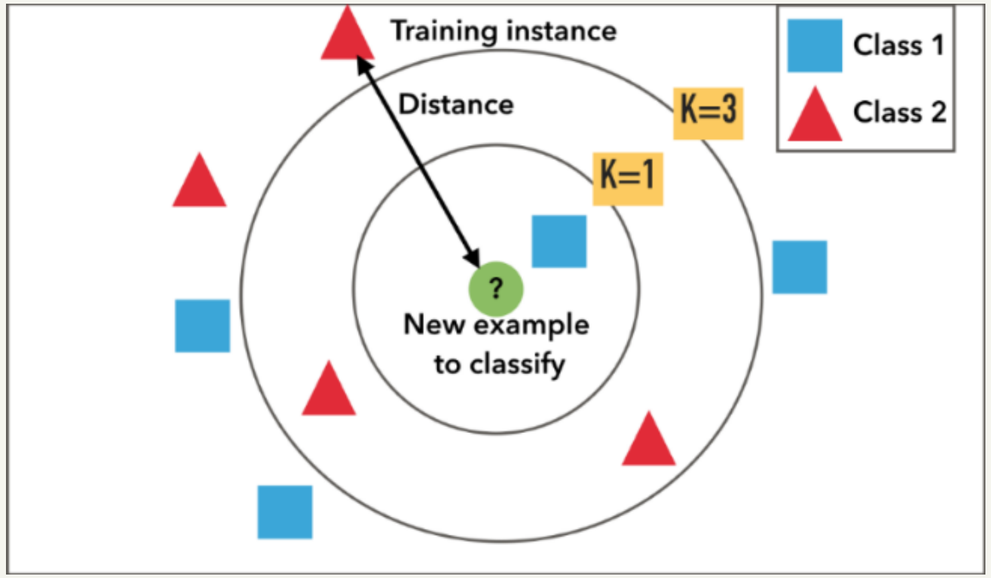
Machine Learning Models

SUPERVISED

K-NEAREST NEIGHBOR

Assumes that objects of similarity are close in distance or proximity.

Finds k training data points that are closest to the data point that requires us to determine a certain classification for.



Machine Learning Models

UNSUPERVISED

K-MEANS CLUSTERING

An iterative algorithm that determines different clusters that could be found in the dataset

The clusters must be distinct non-overlapping subgroups and each data in the dataset must belong to only one group.

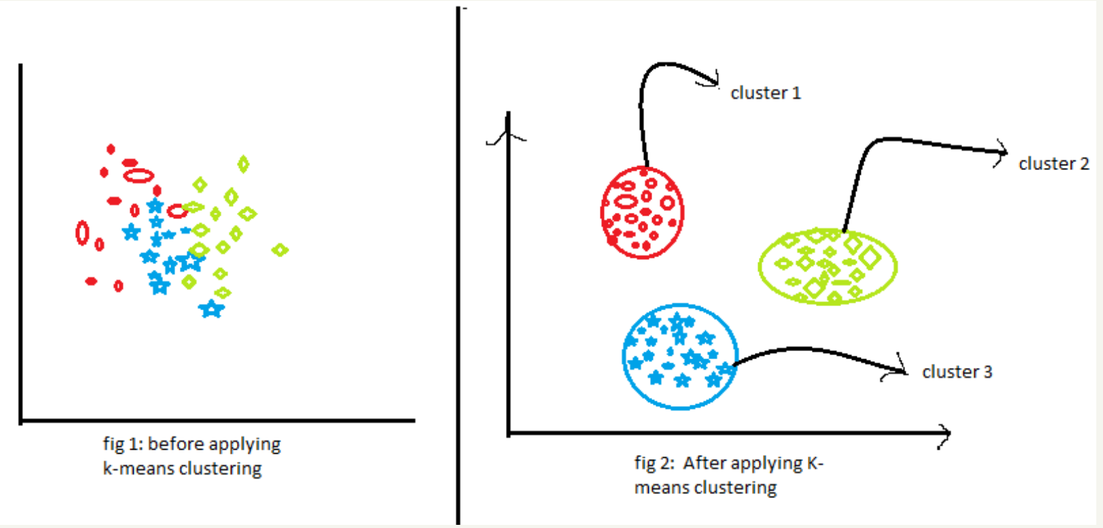


fig 1: before applying k-means clustering

fig 2: After applying K-means clustering

Results of the Study

K-Nearest Neighbor

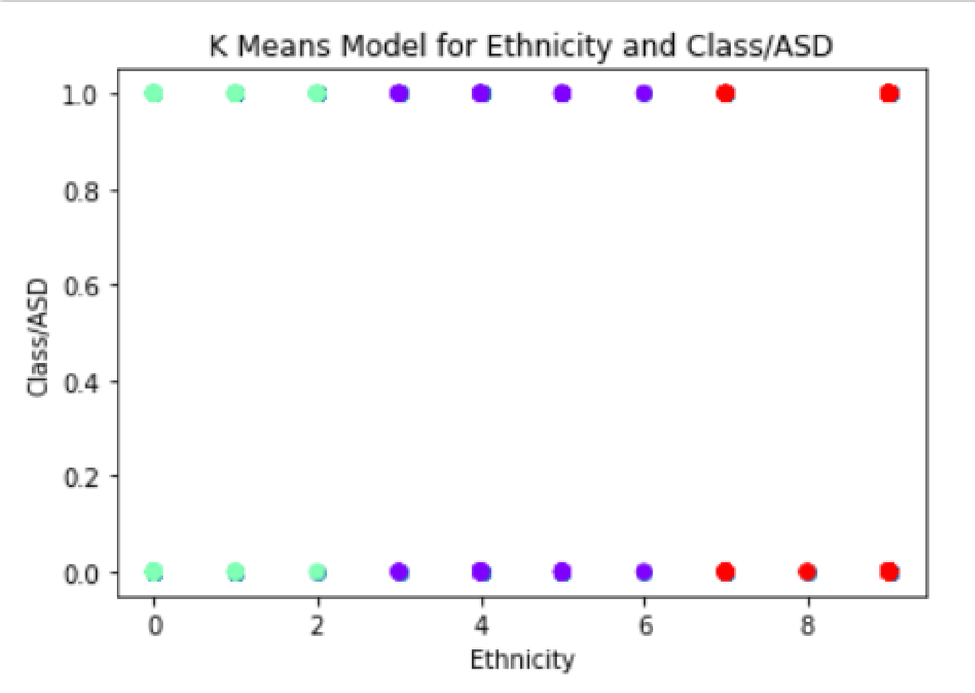
Confusion Matrix: [20 5]
[2 23]

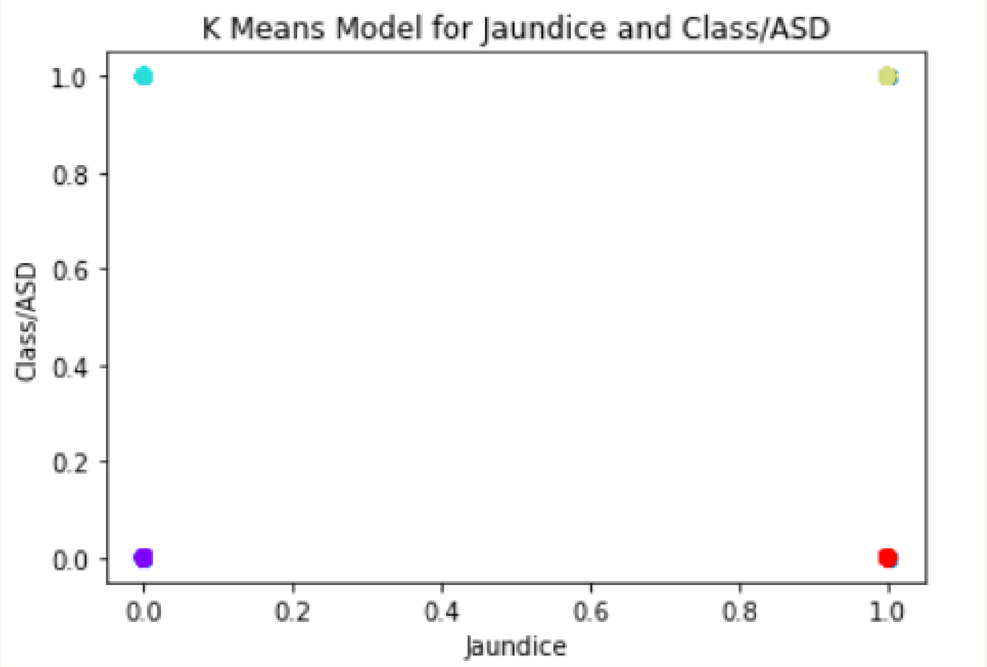
F1 Score: 87%

Accuracy Score: 86%

12 columns were dropped from the base data, i.e. A1 to A10 questions, “age_desc”, and “used_app_before”, for we thought that they are not relevant to the models.

K-Means


K Means Model for Ethnicity and Class/ASD
Class/ASD
Ethnicity
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8


K Means Model for Jaundice and Class/ASD
Class/ASD
Jaundice
0.0
0.2
0.4
0.6
0.8
1.0