# AUTISM SPECTRUM DISORDER (ASD) DETECTION IN toddlers

*A Machine Learning approach*

**Contributors**
**ONUBA CHIBUIKE WINNER***
**OZIGBO CHIDERA DOMINIC****

* onubawinner042@gmail.com
** chideraozigbo@gmail.com

# INTRODUCTION

**Autism spectrum disorder (ASD)** is a complex and diverse condition that affects the development of the brain in various ways. As a result, people with ASD experience challenges and differences in how they communicate, interact, and relate with others, as well as how they express and regulate their behaviors and interests. These behaviors and interests may be limited, repetitive, or unusual, and may vary widely across individuals with ASD

Unfortunately, waiting times for an ASD diagnosis are lengthy and procedures are not cost effective

**What if we use machine learning to predict whether a person suffers from Autism or not?????**

# THE DATASET

The dataset used was developed by Dr Fadi Fayez Thabtah (fadifayez.com) using a mobile app called ASDTests (ASDtests.com) to screen autism in toddlers.

The dataset formed the data collection contains **1054 records** with **18 features** including the outcome (target variable)

- ❑ The dataset does not contain missing values
- ❑ The dataset does not contain outliers

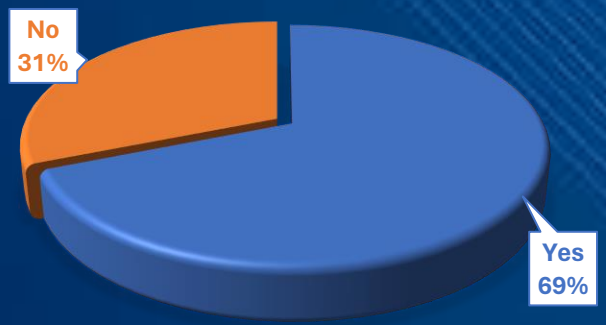
# THE DATASET

**Features on the dataset**

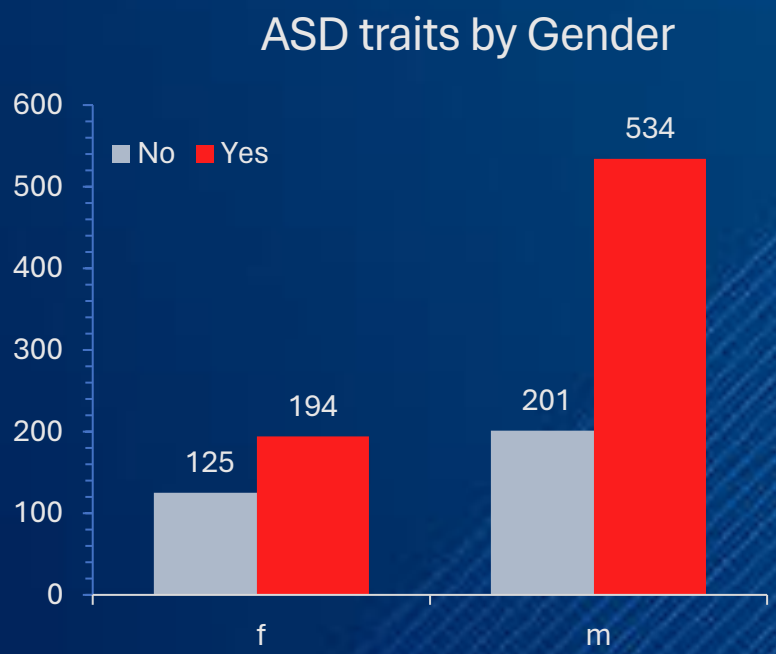
**A1** - Does your child look at you when you call his/her name? **(Binary 0, 1)**
**A2** - How easy is it for you to get eye contact with your child? **(Binary 0, 1)**
**A3** - Does your child point to indicate that s/he wants something? (e.g. a toy that is out of reach) **(Binary 0, 1)**
**A4** - Does your child point to share interest with you? (e.g. pointing at an interesting sight) **(Binary 0, 1)**
**A5** - Does your child pretend? (e.g. care for dolls, talk on a toy phone) **(Binary 0, 1)**
**A6** - Does your child follow where you're looking? **(Binary 0, 1)**
**A7** - If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging them) **(Binary 0, 1)**
**A8** - Would you describe your child's first words as: **(Binary 0, 1)**
**A9** - Does your child use simple gestures? (e.g. wave goodbye) **(Binary 0, 1)**
**A10** - Does your child stare at nothing with no apparent purpose? **(Binary 0, 1)**
**Age (in months) (Number)**
**Score by Q-chat-10 (Number)**
**Sex (String)**
**Ethnicity (String)**
**Born with jaundice (String)**
**Family member with ASD history (String)**
**Who is completing the test (String)**
**Class/ASD Traits (Binary 0, 1)**

# OBSERVATIONS

### ASD OUTCOME DISTRIBUTION



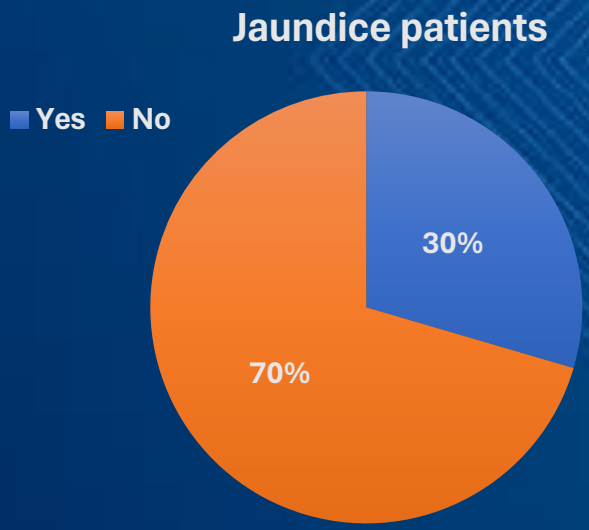

The distribution of the outcome variable is imbalanced with the majority class being the positive outcome.
The imbalance nature of the variable is **SMOTENED** out before modeling building



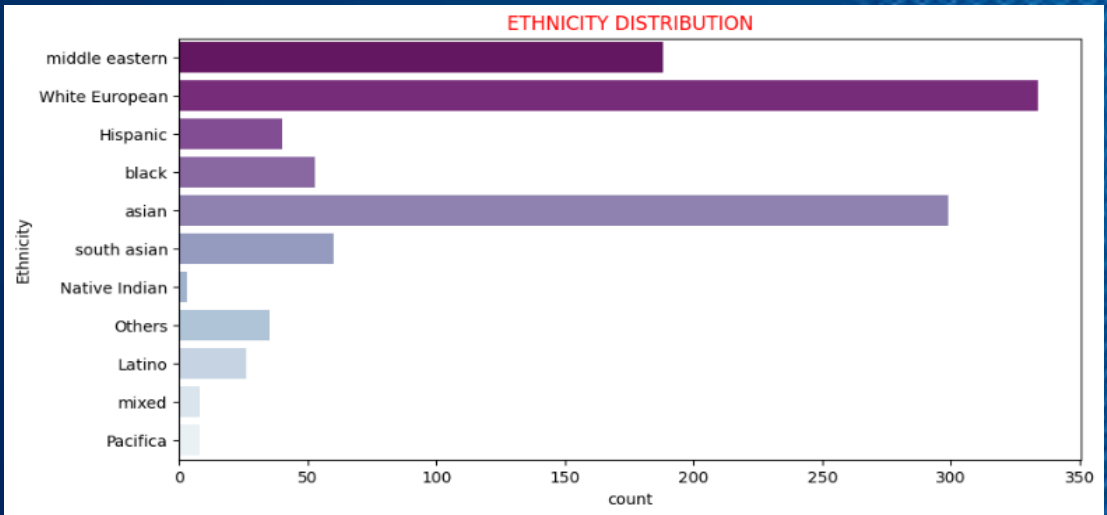

The dataset contained mostly male patients and a high percentage of them were diagnosed with ASD traits

**Check out the EDA notebook for more visualizations**

# OBSERVATIONS




Only about **30%** of the ASD diagnosed patients is also diagnosed with Jaundice signifying little correlation between the data points (ASD and Jaundice)

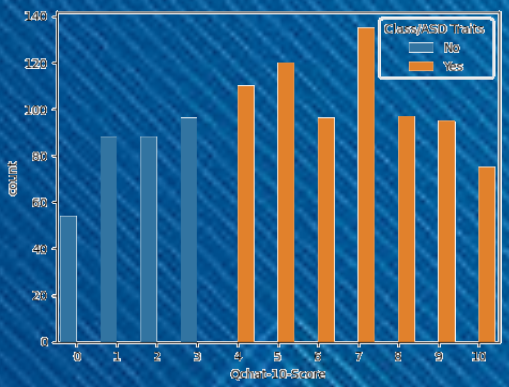



The dataset contains records majorly from white European and Asian children and the least from Native Indian and Pacifica

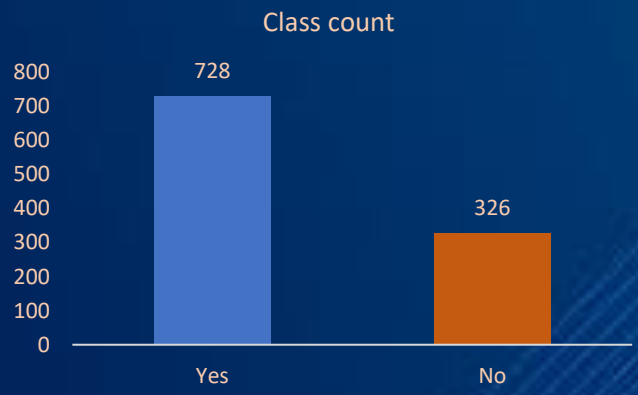
**Check out the EDA notebook for more visualizations**

# THE MODEL

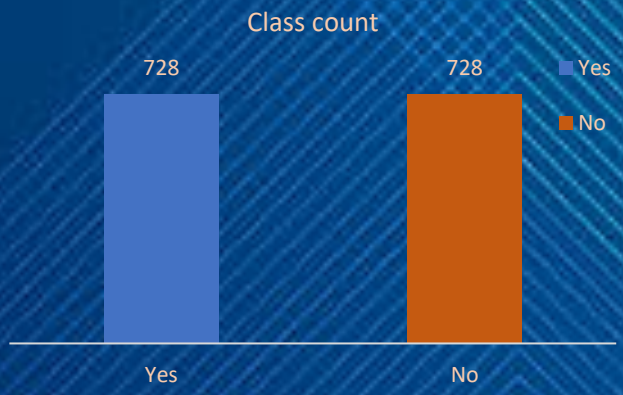
The **Score by Q-chat-10** column in the dataset is a sum of the A1-A10 values. This value has a correlation of over 80% with the target value. Hence the column is dropped to avoid multicollinearity in the model, along side some other columns




Class imbalance in the dataset was addressed using the **SMOTE** technique




**Before balancing**




**After balancing**

# THE MODEL

The dataset was split into a 4:1 ratio for training and testing respectively.

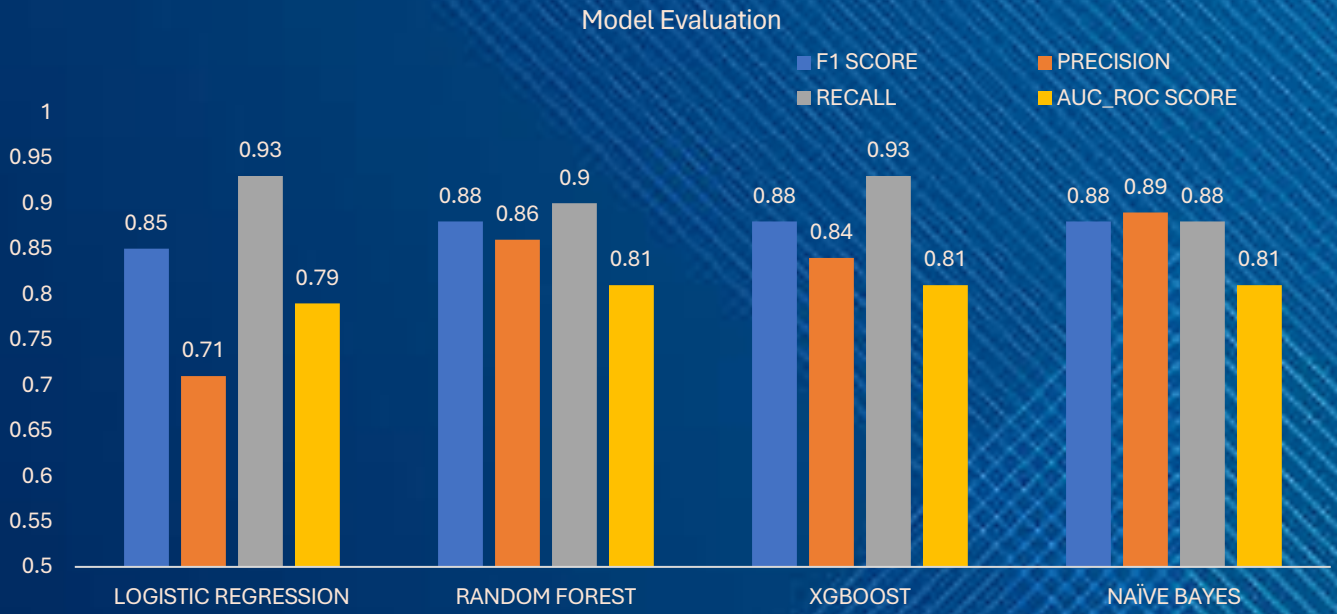
85% of the imbalanced data was used to train the **LOGISTIC REGRESSION**, **RANDOM FOREST CLASSIFIER, XGBOOST CLASSIFIER** and **NAÏVE BAYES** algorithms

After the implementation of the **SMOTENER**, the models were retrained and re-evaluated

The metrics used to evaluate the model are the **F1_Score, Precision, Recall nd AUC_ROC score**

# EVALUATION

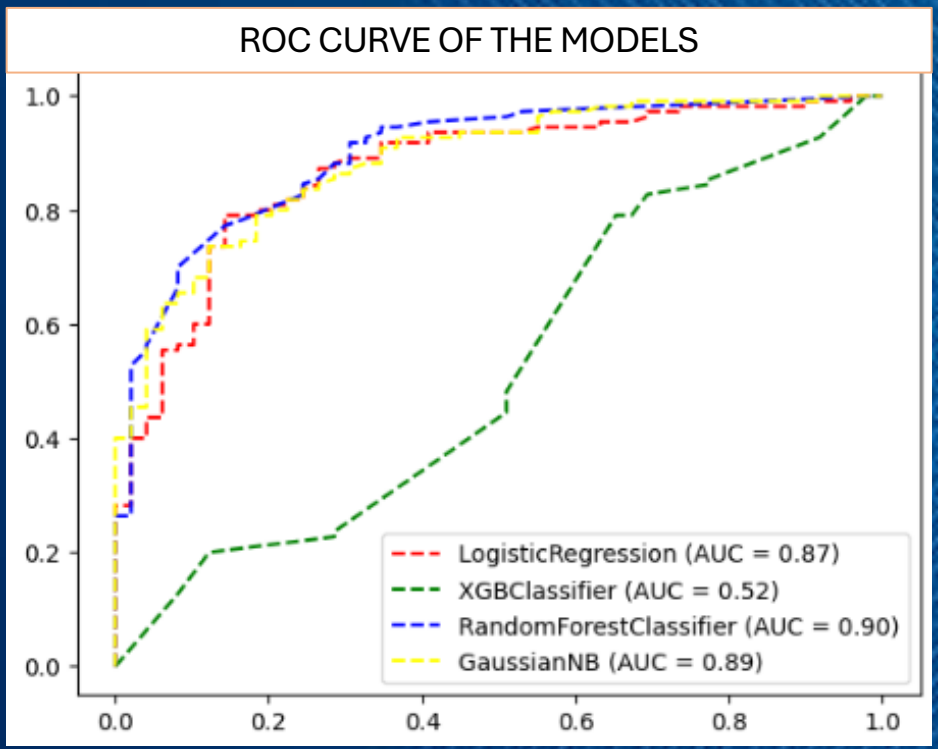
The evaluation report of the models are visualized below




- The logistic regression model performed the least in terms of F1, Precision and AUC score
- On an average, the Naive Bayes classifier performed better than the other models

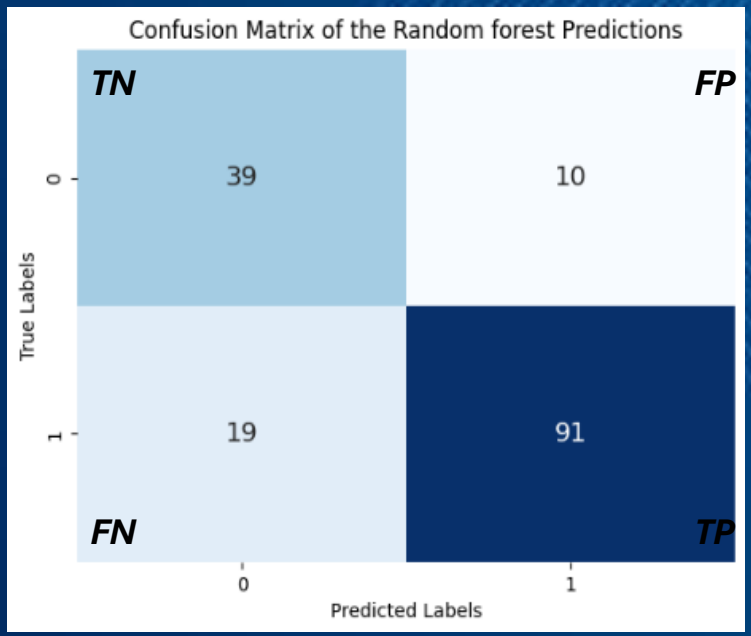
# EVALUATION

When plotting the False Positive Rate against the True Positive Rate of the models, the ROC curve below illustrates the comparison of the models' efficiency.




After looking at the ROC curve for different models, we observe that the **Random Forest** performed the best, showing better results in the binary classification task

# THE RESULT




$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN}$$

*The accuracy is the proportion of correct predictions out of the total predictions. As shown in the matrix, the model has an accuracy of 0.82.*

*This means that the model can correctly identify **82%** of the cases, and has a high ability to detect both ASD and non-ASD cases. The model also has a low rate of false positives and false negatives, which indicates that it is **reliable** and **robust**.*

# REFERENCES


- Vakadkar, K., Purkayastha, D., & Krishnan, D. (2021). Detection of autism spectrum disorder in children using machine learning techniques. SN Computer Science, 2, 386

- Hossain, M. D., Kabir, M. A., Anwar, A., & Islam, M. Z. (2021). Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets. Health Information Science and Systems, 9, 17

- Oyebode, O., Oyebode, T., & Oyebode, F. (2020). Using machine learning optimization to predict autism in toddlers. In Proceedings of the International Conference on Industrial Engineering and Operations Management (pp. 1-10)

Real World AI Project

Date: 18 March 2023

# Certificate of Completion

This award is given to

# Onuba Chibuike Winner

## FOR THE HIGH COLLABORATION IN THE PROJECT:

***Predicting autism in toddlers using machine learning***

**IMPACT** This certificate recognizes the valuable contributions made as a collaborator on the research project that developed a machine learning model for predicting autism in toddlers. The resulting research paper showcases the potential of this groundbreaking approach for early detection, enabling timely intervention and positively impacting the lives of countless children affected by autism spectrum disorder.

| | | |
|---|---|---|
| **Vidura Wijekoon** | **Varuna Bamunusinghe** | **Hiran Wijesinghe** |
| Sri Lanka Chapter Lead | Sri Lanka Chapter Lead | Sri Lanka Chapter Lead |

mdena LOCAL CHAPTERS