# ARTIFICIAL INTELIGENCE

## SUPERVISED LEARNING

In this project we will study and create a program to address detection of autism in toddler

Autism Dataset for Toddlers

https://www.kaggle.com/datasets/vaishnavisirigiri/autism-dataset-for-toddlers

# DOCUMENTATION REFERENCES

## Documentation

Scikit Documentation

SMOTE Documentation

## Resources from Moodle

Pandas Cheat Sheets

Machine_learning_Python_Cheat_Sheets

Scikit learn Edureka Cheat Sheet

Scikit Learn Cheat Sheet

# TOOLS USED

## Python

Pandas
Matplotlib
Seaborn
Numpy
Sklearn
Imblearn

# IMPLEMENTATION

Initially, we utilised Pandas to construct the dataframe.
Next, employing Matplotlib, we conducted an analysis to eliminate redundant features right from the start.
Non-numerical features deemed pertinent were encoded.
We generated a correlation matrix to identify highly correlated features.
Due to a discrepancy in result counts, we opted to employ SMOTE for oversampling.
Subsequently, we trained our model using four different classifiers.

Decision Tree Classifier

Multi-layer Perceptron Classifier

k-Nearest Neighbors Classifier

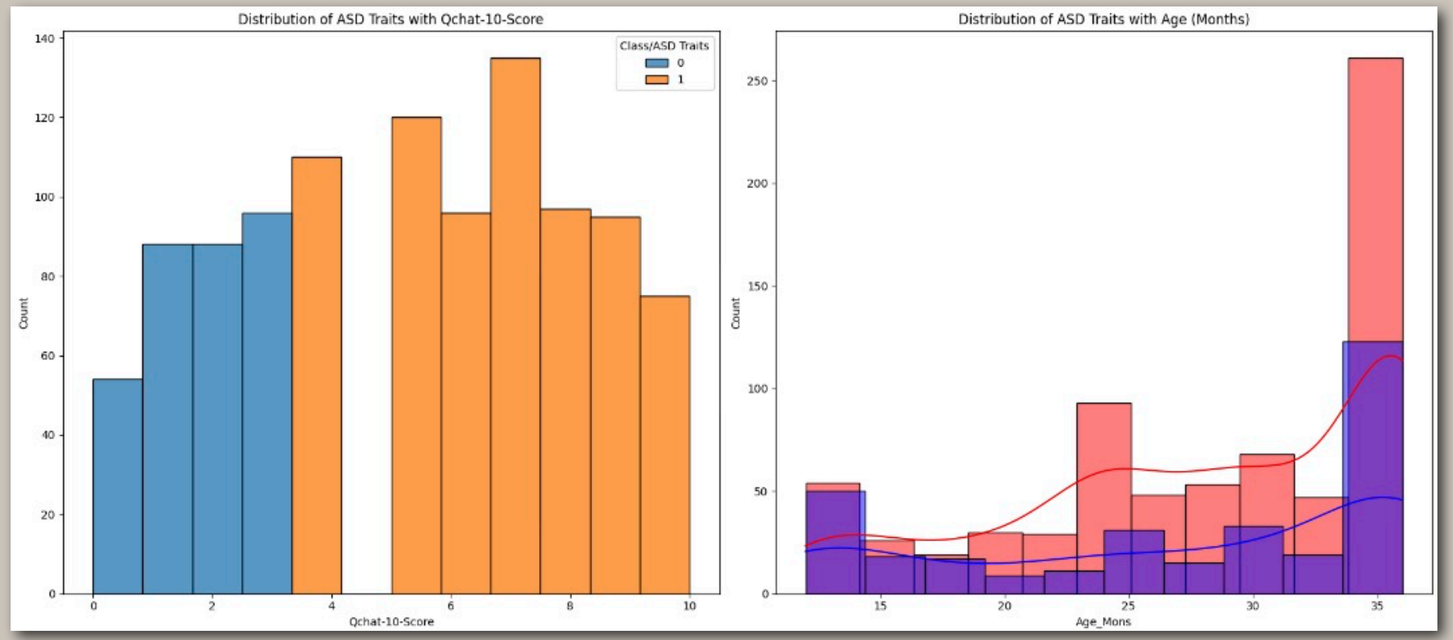
Support Vector Machine Classifier

# DATA PREPOCESSING

We first created a dataframe using Pandas.

| | Case_No | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 | Age_Mons | Qchat-10-Score | Sex | Ethnicity | Jaundice | Family_mem_with_ASD | Who completed the test | Class/ASD Traits |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 28 | 3 | f | middle eastern | yes | no | family member | No |
| 1 | 2 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 36 | 4 | m | White European | yes | no | family member | Yes |
| 2 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 36 | 4 | m | middle eastern | yes | no | family member | Yes |
| 3 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 24 | 10 | m | Hispanic | no | no | family member | Yes |
| 4 | 5 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 20 | 9 | f | White European | no | yes | family member | Yes |

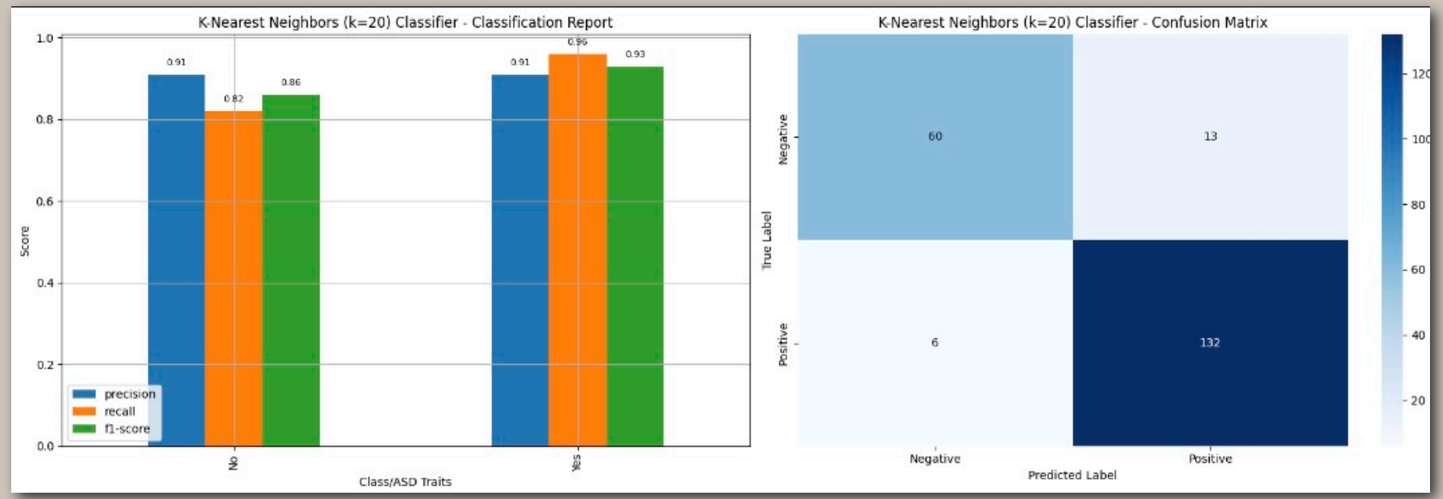
Since some of the features are non numerical, we must encode them and we can even remove some features we do not need.



# MODELS

## Decision Tree Classifier

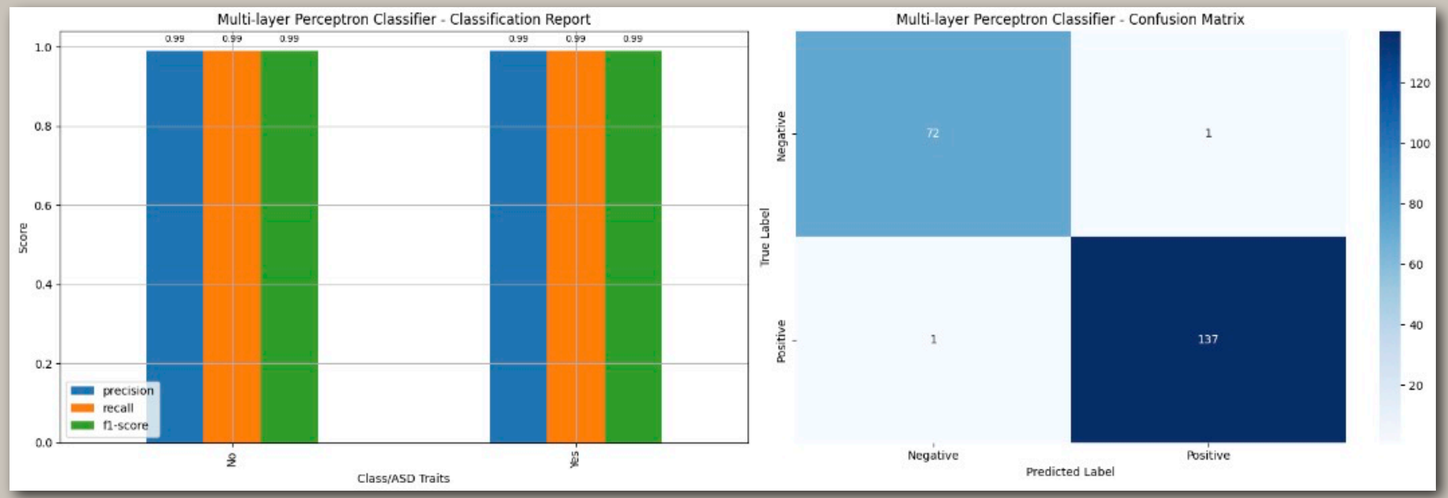

## k-Nearest Neighbours Classifier



# MODELS

## Support Vector Machine Classifier


## Multi-layer Perceptron classifier



# MODELS COMPARISON

## Good Performers: Decision Tree and K-Nearest Neighbors (KNN):

Both the Decision Tree Classifier and the K-Nearest Neighbors (KNN) Classifier demonstrated solid performance in this analysis. The Decision Tree Classifier offers interpretability and ease of use, making it a suitable choice for datasets with binary features. Additionally, KNN achieved an optimal balance between bias and variance, resulting in accurate predictions. These models are effective in scenarios where the decision boundary is relatively simple and well-defined.

## Poor Performer: Support Vector Machine (SVM):

The Support Vector Machine (SVM) Classifier performed relatively poorly compared to other models. This could be attributed to its sensitivity to dataset characteristics, particularly when the data is not linearly separable. SVMs struggle to handle complex datasets with nonlinear decision boundaries, resulting in suboptimal performance. Additionally, SVMs may require careful tuning of hyperparameters and kernel functions to achieve better results.

## Top Performers: Random Forest and MLPClassifier:

The Random Forest Classifier and the MLPClassifier (Neural Network) emerged as the top performers in this analysis. The Random Forest Classifier demonstrated robustness and generalization capability, effectively capturing intricate patterns in the data. Its ensemble learning approach reduces overfitting and improves performance, making it suitable for complex datasets. Similarly, the MLPClassifier showcased high accuracy and the ability to capture complex relationships in the data. Neural networks excel at learning nonlinear relationships, making them well-suited for datasets with binary features.