Autism Diagnosis

Presented By:

Khulood Nasher

Define the Problem

Autism spectrum disorder (ASD) is a type of neurological and developmental disorder that begins early in childhood and continues throughout a person's life. In this project, Dr. Fayez Thabtah collected data from 1054 toddlers whose ages were less than three years old. The Doctor asked their parents yes/no questions. These questions were behavioral questions and the number of yes were added up and have been recorded in the column called Q chat 10 question. If this column has more than 3 yes, toddlers will be classified as positive ASD.

Our role in this project is to use machine learning techniques to diagnose ASD based on the Dr. Thabtah's data.

Important Questions

1-What is good algorithm to diagnose autism?

2- How can I improve my models' accuracy?

3- What are the most important features in diagnosing autism among children?

Methodology

Obtain Data
1054 Toddlers
18 features
Features:10 yes/no questions
Ethnicity, gender, age in months Jaundice (yes/no) Family history of autosim (yes/no) family member
 Fill out questionnaire

Preprocessing Data
- cleaning data: Removing unwanted columns.
- Encoding categorical columns - Explored columns: Visualization of positive versus negative Heatmap correlation Barplot of age .ethnicity, gender,jaundice, chat 10 questionnaire

Modeling
data were split 60% train and 40% test

SVC, SGD,Logistic Regression,Random Forest, NB, Adaboost, XGBoost, Neural Network,
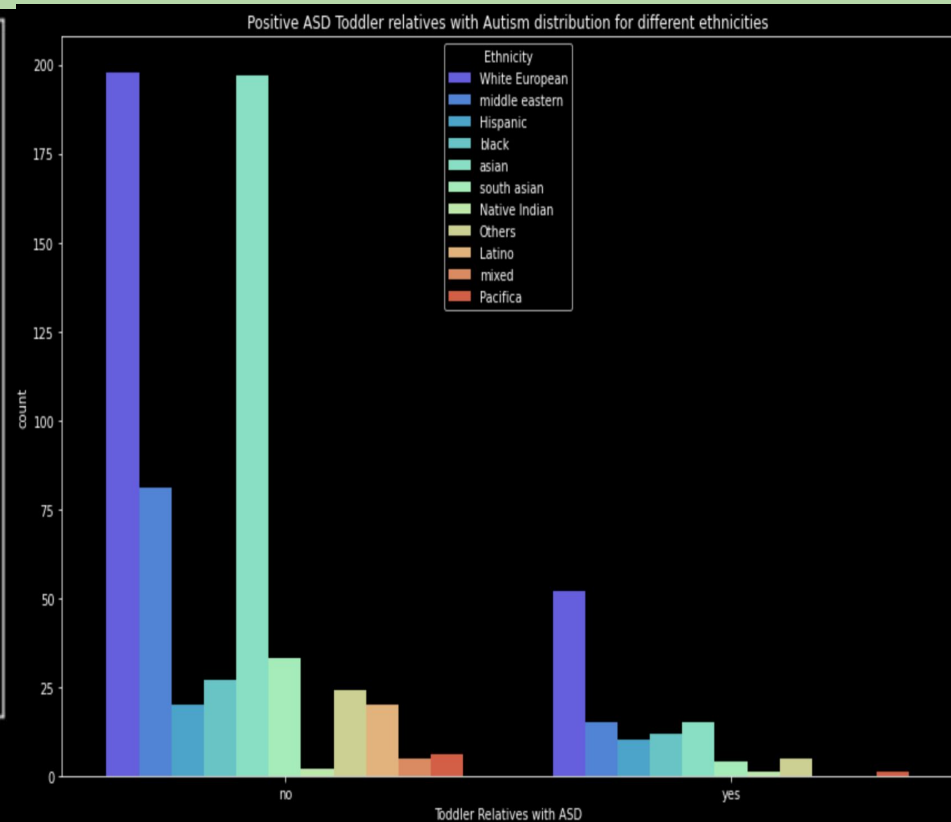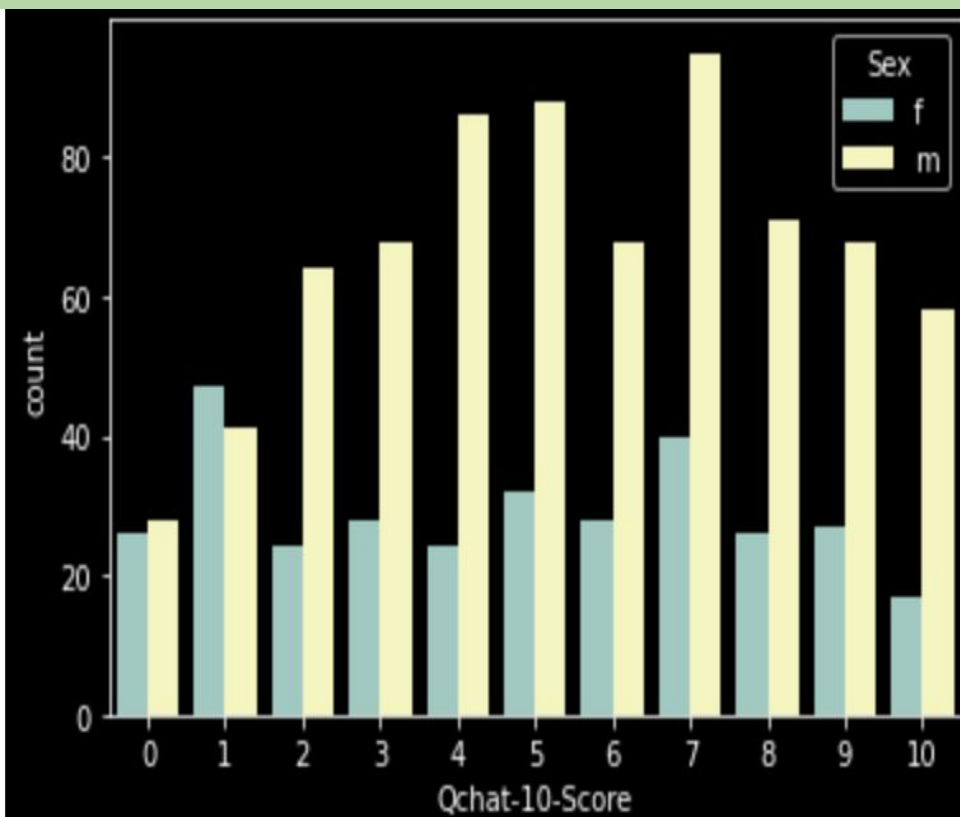
Accuracy Metrics
Testing Accuracy :99% in Logistic Regression, XGBoost, and neural network
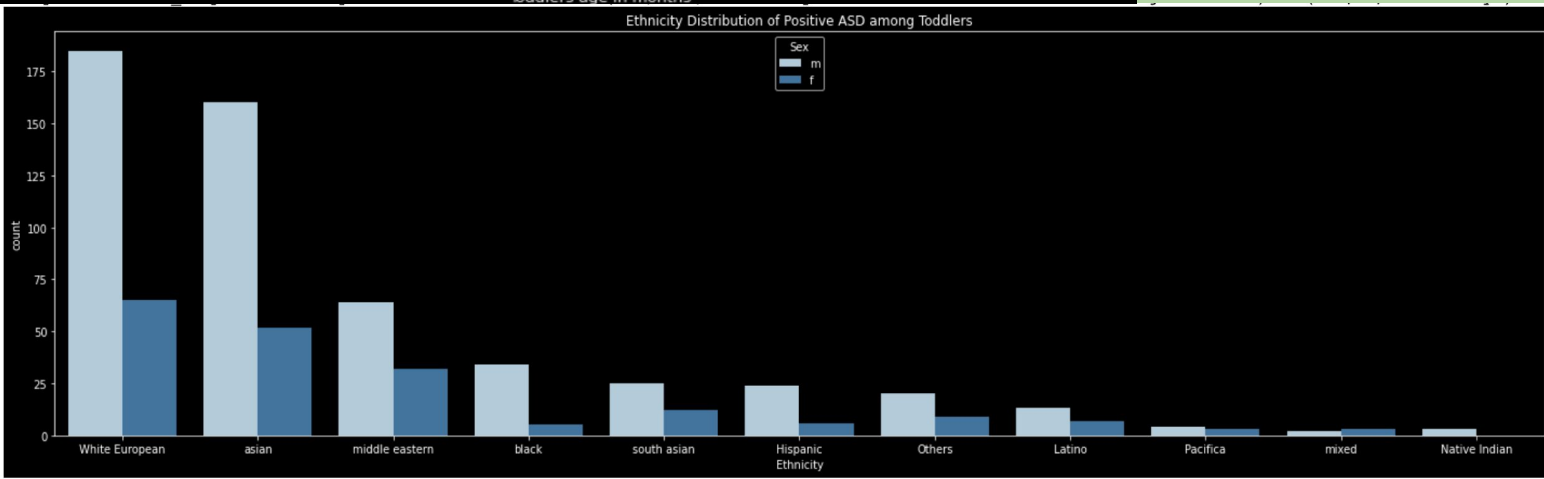
Metrics: Confusion Matrix F1, Recall, Precision

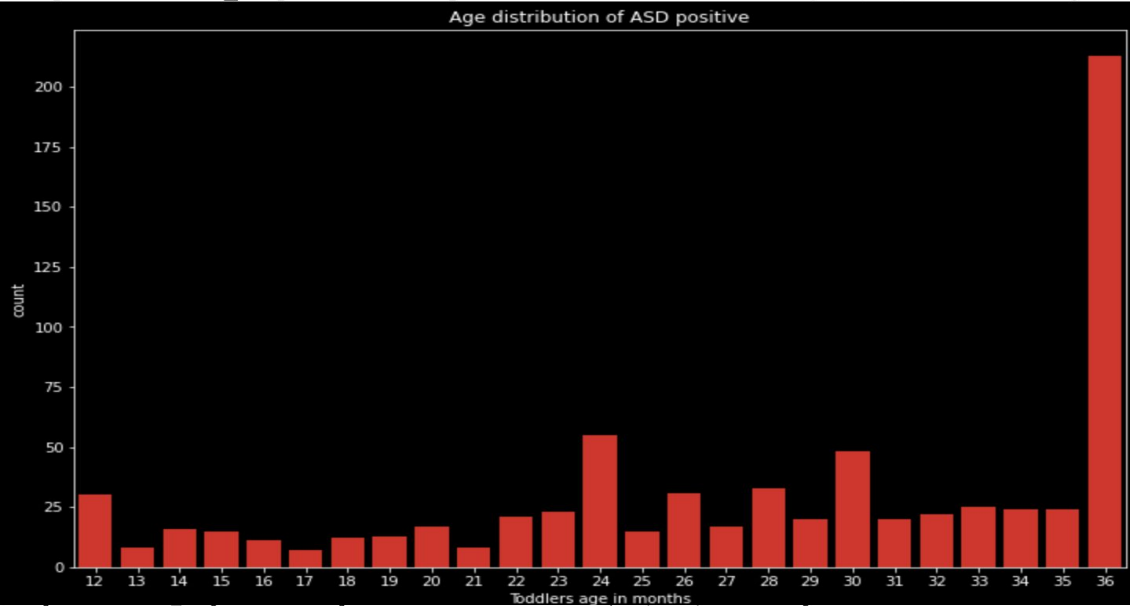|		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

Deploy The Model

Explore Data

Age distribution of ASD positive

Ethnicity Distribution of Positive ASD among Toddlers

Explore Data

Modeling Results

Out[48]:

	Classifier	test_Accuracy
0	LogisticRegression	1
1	LinearDiscriminantAnalysis	0.96
2	KNeighborsClassifier	0.91
3	DecisionTreeClassifier	0.91
4	GaussianNB	0.94
5	SVC_beforegrid	0.78
6	RandomForest_beforegrid	0.64
7	XGBClassifier	0.99
8	GradientBoosting	0.64
9	AdaBoosting	0.49
10	SVC_aftergrid	1
11	RandomForest_aftergrid	0.96
12	Neuralnetwork_SKLearn Accuracy	0.99
13	Neuralnetwork_Keras	0.95

LogisticRegression, XGBClassifier, and Neuralnetwork_SKLearn are the most accurate classifiers.
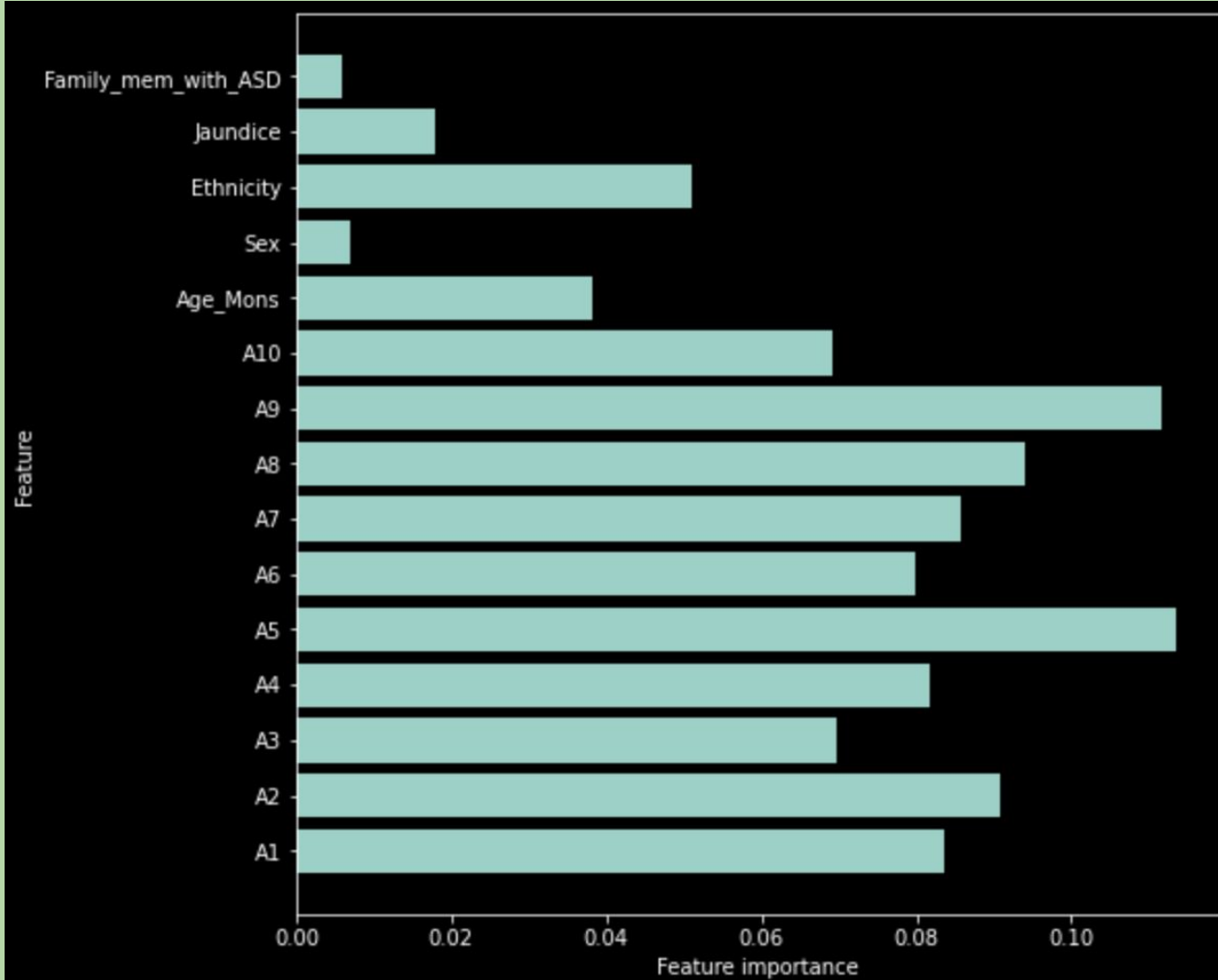
Confusion Matrix
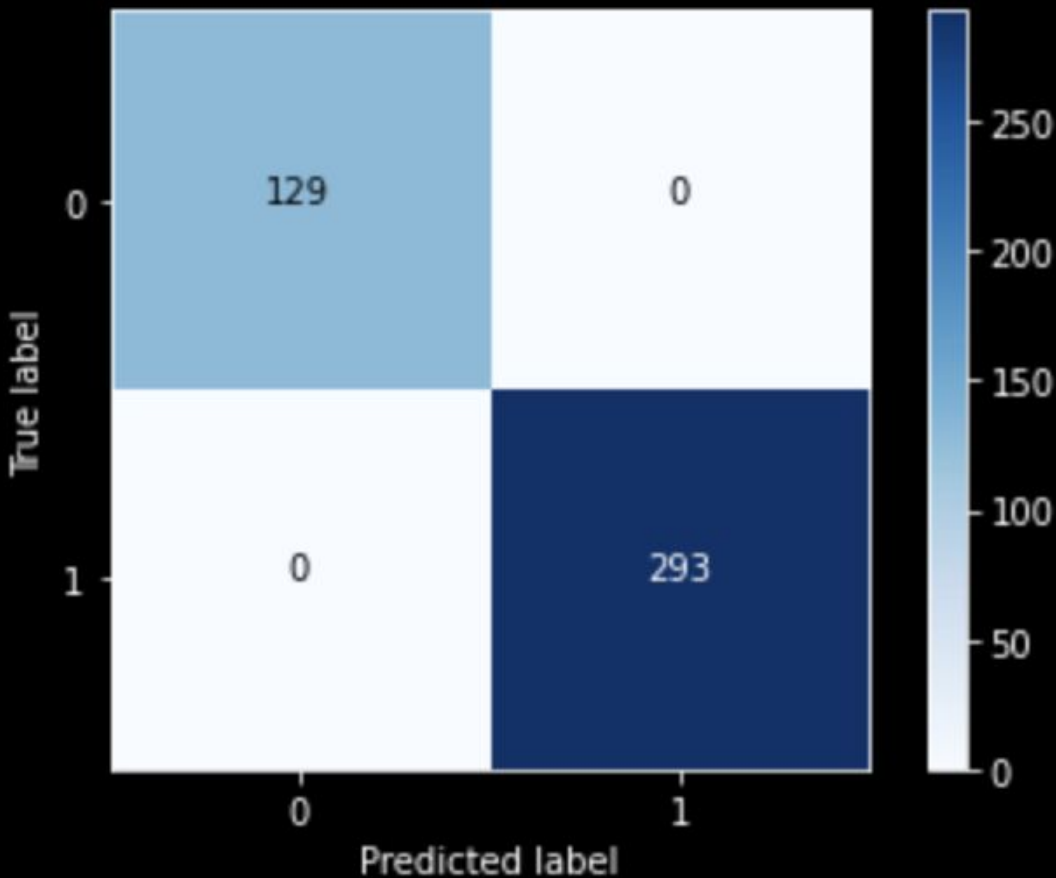

```
test accuracy: 1.0
train accuracy: 1.0
```

Metrics

False-Negative is the most important number in a health center where the model predicts the patient is negative/healthy while the truth is that the patient is positive/sick.

Our model misclassified 0 patients.

Important Features

Conclusion

Models

- Logistic Regression, XG Boosting and Neural Network are good models in our autism project with it small sample size.

Important Features

- A9 is the most important features, other important features are A6, A5, and A8. Jandice and the type of gender'sex' are less important features. Genes in this study which appears through the column of family member with ASD is not an important feature
- Males are more positive to autism than females.
- Ages close to 36 months which is three years old show more positive autism.
- White European, Asian and middle eastern are the ethnicities that showed an increase in autism cases

Recommendation

- Apply Logistic Regression, XG Boosting or neural network using sklearn for high accuracy model to diagnose Autism.
- Diagnose deeply and early autism disease among particular ethnicity particularly, White European, Asian and Middle Eastern.
- Try to find more state of art in tracing autism symptoms among toddlers in ages less than three years.

Future Work

- Reach out the same clinic and get more data either with number of diagnosed patients or with more features used for diagnosis.
- Try to ask more numerical-based questions to the children
- Search other data on Autism around the world and do the same approach used in this project.
- Apply the same models on adults
- Model with image data of cases diagnosed with autism using X-ray

Thank you

Any questions?