Autism Prediction using Machine Learning

Aim:

The aim of this project is to leverage machine learning techniques to predict whether a person suffers from Autism Spectrum Disorder (ASD) or not. ASD is a complex neurological disorder impacting social interaction, communication, and behavior. Since conventional diagnostic methods for ASD are limited, this project explores the potential of machine learning algorithms to predict ASD based on various features.

Description of the Data:

The dataset used in this project contains information on individuals, including demographic details, behavioral indicators, and clinical scores. Key features include age, gender, ethnicity, clinical scores from A1 to A10, and other behavioral attributes. The target variable is the presence or absence of ASD.

Analysis:

- **Data Cleaning:** The initial step involved cleaning the dataset, addressing null values, and standardizing categorical variables.
- **Exploratory Data Analysis (EDA):** EDA was performed to understand data distribution, identify patterns, and visualize relationships between variables. Key insights included identifying data imbalances, skewness, and correlations between features.
- **Feature Engineering:** Additional features were derived from existing ones to enhance model performance. Age groups were created, and clinical scores were aggregated to provide a comprehensive view of ASD likelihood.
- **Model Training:** Various machine learning models, including Logistic Regression, XGBoost, and Support Vector Classifier (SVC), were trained on the dataset. Techniques such as Random Over Sampling were used to address data imbalance, and feature scaling was applied for stable training.
- **Model Evaluation:** The trained models were evaluated based on training and validation accuracy. Confusion matrices were plotted to visualize model performance.

Conclusion:

The machine learning models achieved an accuracy ranging from 80% to 85% in predicting ASD, demonstrating promising results in the absence of definitive diagnostic methods. Logistic Regression and SVC classifiers exhibited better performance on the validation data, highlighting their potential for real-world applications in ASD prediction. This project showcases the utility of

machine learning in addressing complex neurological disorders and underscores the importance of data-driven approaches in healthcare.