Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

NAME: **Shubhjyot Singh Chadha** SAP ID: **60009220197**

COURSE NAME: **Machine Learning - I Laboratory** BATCH: **D2-2**

Mini Project

Task 4

(Model)

Title of Your Project: Autism Prediction/Detection

Aim of the Project:

Autism, or autism spectrum disorder (ASD), encompasses a wide range of conditions characterized by difficulties in social skills, repetitive behaviors, and communication. It's influenced by both genetic and environmental factors. Each individual with autism presents a unique set of strengths and challenges, ranging from highly skilled to severely affected. Early intervention can significantly improve learning, communication, and social skills, as well as underlying brain development, but the diagnostic process often takes years.

The primary objective of the project is to make a predictive model to determine whether one suffers from autism or not. This model will assist doctors and individuals to get an initial diagnosis and detect autism so the individual suffering from it can get the help they need.

Data Description:

This dataset is composed of survey results for more than 700 people who filled an app form. There are labels portraying whether the person received a diagnosis of autism, allowing machine learning models to predict the likelihood of having autism, therefore allowing healthcare professionals prioritize their resources

The dataset comprises several attributes which have been described below:

- ID: ID of the patient (Numerical)
- A1_Score to A10_Score: Score based on Autism Spectrum Quotient (AQ) 10 item screening tool (Numerical)
- Age: Age of patient in years (Numerical)

- Gender:Gender of the patient (Male/Female) (Categorical)
- Jaundice: Whether the patient had jaundice at the time of birth (Yes/No) (Categorical)
- Autism: Whether an immediate family member has been diagnosed with autism (Yes/No) (Categorical)
- Ethnicity: Ethnicity of the patient (Categorical)
- Contry_of_res: Country of residence of the patient (Category)
- used_app_before: Whether the patient has undergone a screening test before (Yes/No) (Categorical)
- Result: Score for AQ1-10 screening test (Numerical)
- Age_desc: Age of the patient (Numerical)
- Relation: Relation of patient who completed the test (Categorical)
- Class/ASD: Classified result as 0 or 1. Here 0 represents No and 1 represents Yes.

Data Preprocessing:

Dropping Unnecessary Variables:

- Dropped age_desc column as we found it was not useful to us after analysis of the data.

Encoding Categorical Variables:

- Categorical variables are converted into a format that can be provided to ML models. Techniques such as one-hot encoding or label encoding are used depending on the algorithm requirements.
- Gender, Autism, Jaundice, Used_app_before are converted to numerical values of 1/0 as they are bivariate categorical values.
- Ethnicity and Country_of_res variables are converted to numerical values and have been scaled to a range, between 0 and 1.

Scaling Features:

- Numerical features are scaled using standardization or normalization to ensure that no variable dominates another due to its scale, improving the stability and performance of the learning algorithm.

Data Modeling:

It is a classification problem hence multiple classifiers were implemented to check which model gives us better accuracy.

Several machine learning models were chosen for training based on their suitability for classification tasks and their ability to handle the dataset's characteristics.

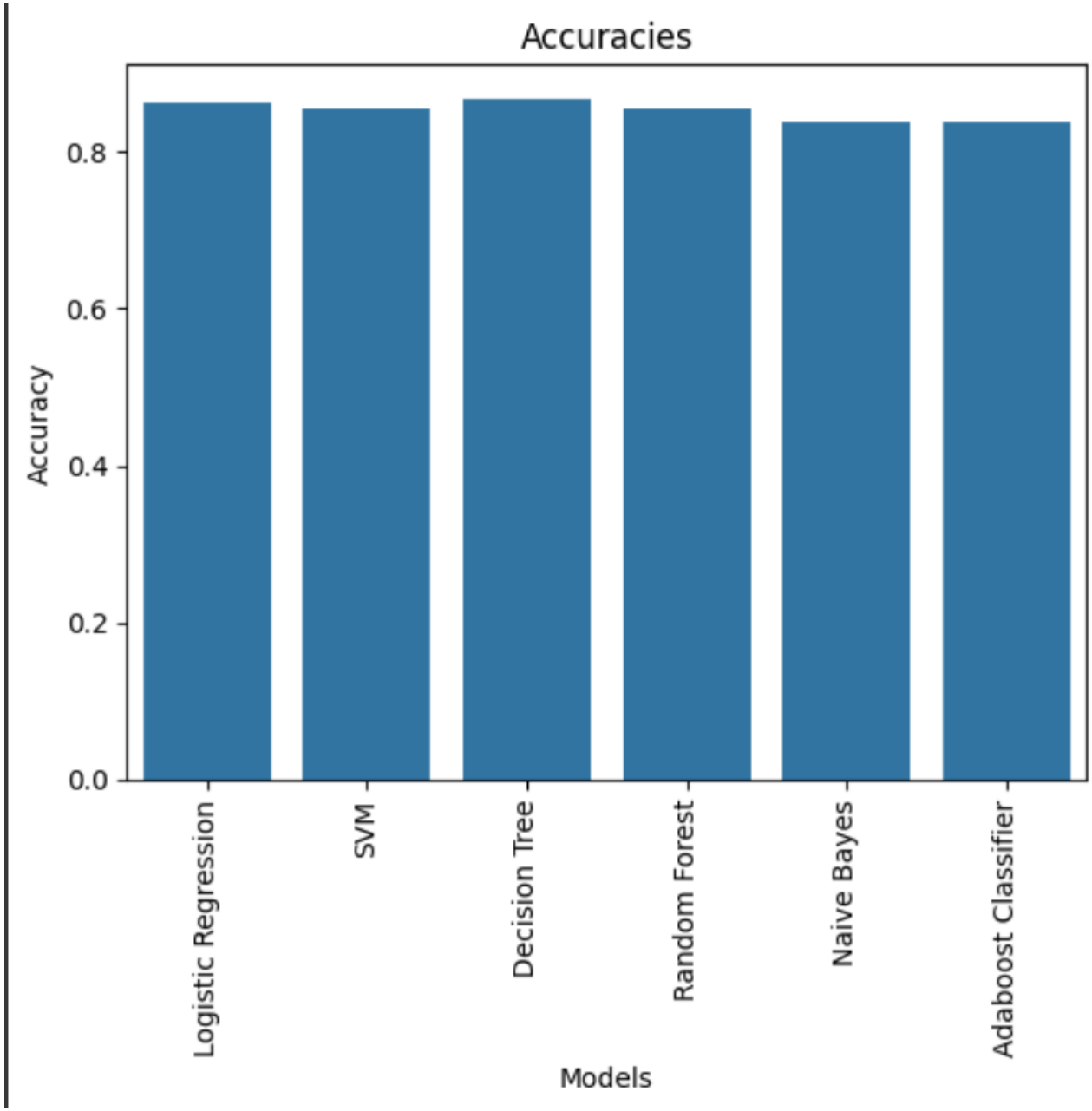
The selected models included:

- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree
- Random Forest
- Naive Bayes
- Adaboost Classifier

Each model was trained on the preprocessed dataset using appropriate libraries such as scikit-learn in Python.The dataset was split into training and testing sets to evaluate the models' performance on unseen data.

Performance Evaluation:

Model performances were compared using metrics such as accuracy, precision, recall, and ROC area under the curve. The models' ability to accurately predict Autism was further analyzed using confusion matrices. The Decision Tree model performed the best in terms of accuracy and overall balance between sensitivity and specificity, making it the model of choice for Autism Prediction.


Accuracies
Accuracy
0.0
0.2
0.4
0.6
0.8
Logistic Regression
SVM
Decision Tree
Random Forest
Naive Bayes
Adaboost Classifier
Models