A Comparative Analysis of Machine Learning models in the Classification Techniques in the Prediction of Autism in Children

Ebierimunu ABULE
LCU/PG/002902

Being a MSc Viva Presentation to the Department of Computer Science, Faculty of Natural & Applied Sciences, Lead City University, Ibadan, Oyo State, Nigeria

In Partial Fulfillment of the Requirements for the Award of Master of Science Degree (MSc) in Computer Science

2025

Certification

This is to certify that this study titled "A Comparative Analysis of Machine Learning models in the Classification Techniques in the Prediction of Autism in Children

" was conducted by Ebierimunu ABULE with Matric No. LCU/PG/002902 for the award of Master of Science Degree (MSc) in Computer Science in the Department of Computer Sciences, Faculty of Natural and Applied Sciences, Lead City University Ibadan, Oyo State, Nigeria under my supervision, and that this has not been previously submitted.

……..……..

Dr. Ayoade Akintayo

Supervisor

…...…..……

Date

……..……

Dr. V.B Oyekunle

Head of Department

……...……

Date

Dedication

This thesis is dedicated to Almighty God.

Acknowledgement

The Researcher wish to acknowledge and express my sincere gratitude to the Management of Lead City University Ibadan, Oyo State for their support, aligned academic facilities, and ambient environment provided to facilitate the achievement of this programme.

I also in no small measure appreciated the contribution and effort of my able supervisor Dr. Ayoade Akintayo for his unrelenting support and constructive criticisms in ensuring that this project is actualized. Thank you, sir, for your fatherly love. May Almighty God continue to bless you and may he rescue you in point of need.

I cannot help but appreciate the effort and support of the HOD Dr. Sakpere Wilson. I equally appreciate the other members of lecturers in the department of computer science department, Lead City University, Ibadan for their impacts in my academic journey.

Also, deepest appreciation goes to my parents Mr. and Mrs. Lawrence Abule. To my guardian Mr and Mrs. Capt. Wilson Theophilus, to my siblings Engr K. A. Abule, Opuoyere L. Abule, Pharm Ayibanua Theophilus and host of others for their support and friendship role displayed. My God Almighty direct our part and show us the way to success.

Although, all the aforementioned Institutions and persons mentioned helped in one way or the other, meanwhile all errors in the research report if found are solely mind.

Abstract

Autism Spectrum Disorder (ASD) is a mild cognitive impairment known to affect a person's language, communication, thought process and social behaviour. Some recent advances in machine learning made it possible to predict ASD using behavioural,demographic, and medical data, but the choice of optimal algorithms and feature selection techniques remains an open research question. This study used the ASD for children dataset obtained from the University of California, Irvine repository to investigate the performance of the selected machine learning models through a comparative analysis. Eight nonparametric models were compared in total, with four models SVM, KNN, GNB, and MLP being base learners, and the other four ensemble methods RF, XGBoost, Bootstrap Aggregating (Bagging), and Stacking. Hyperparameters of the models were autotuned, and GridSearchCV was employed to choose the model with the best combination of parameters. All base models performed brilliantly, with SVM being the best and most precise in this experiment with a result of 1, amongst all classifiers even before hyperparameter tuning. After hyperparameter tuning, SVM and MLP ranked highest with all metrics returning scores of 100%. KNN returned 96.2% accuracy and recall and GNB trailed at 95.26% for both metrics for the ensemble methods, XGB and Bagging with SVM performed best at 100% for all metrics. RF produced an accuracy and recall of 98.1% while Bagging with KNN and GNB had both metrics at 95.2%. The classification performance of Bagging with MLP and the stacking classifier were equivalent across all metrics though, producing accuracy and recall of 99.5%. Based on the findings, it is recommended that thorough feature selection should be conducted to eliminate features that could lead to overfitting, such as "Q_Chat_10" in this study cross-validation technique was used to assess the generalizability of models to unseen data.

Keywords: Accuracy, Classifier, Defaults, Financial, Models, Predicting, Validation

Word Count: 300

Table of Contents

Chapter Two: Literature Review

List of Tables

Table **Title** **Page**

List of Figures

Figure	Title	Page

List of Acronyms

Abbreviation	Meaning
ABIDE	Autism Brain Imaging Data Exchange
ADOS	Autism Diagnostic Observation Scale

AI	Artificial Intelligence
ASD	Autism Spectrum Disorder
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
DSM-V	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
DL	Deep Learning
EEG	Electroencephalogram
fMRI	Functional Magnetic Resonance Imaging
GNB	Gaussian Naive Bayes
KNN	K-Nearest Neighbours
MLP	Multilayer Perceptron
MRI	Magnetic Resonance Imaging
NHS	National Health Service
RBF	Radial Basis Function
RF	Random Forest
ROC	Receiver Operator Characteristic
SVM	Support Vector Machine
UCI	University of California, Irvine
XGBoost	Extreme Gradient Boosting