A FIELD PROJECT REPORT

on

**“NEURO DIVERSE PERSPECTIVE USING MACHINE LEARNING”**

**Submitted**

by

|  |  |
| --- | --- |
| 221FA04596  Ch. Chaitanya | 221FA04672  Vamsi |
| 221FA04683 221FA04691  R. Avinash S. Hemanth | |
|  | |

**Under the guidance of**

*DR. Deva Kumar*

*Designation*



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH Deemed to be UNIVERSITY**

**Vadlamudi, Guntur.**

**ANDHRA PRADESH, INDIA, PIN-522213.**



**CERTIFICATE**

This is to certify that the Field Project entitled **“NEURO DIVERSE PERSPECTIVE USING MACHINE LEARNING”** that is being submitted by 221FA04596 (Ch.Chaitanya), 221FA04672(Vamsi), 221FA04683(R.Avinash)**&** 221FA04691 (S.Hemanth) for partial fulfilment of Field Project is a bonafide work carried out under the supervision of Dr.Deva kumar , Assistant Professor, Department of CSE.

|  |  |  |
| --- | --- | --- |
| Guide name& Signature |  | Dr.K.V. Krishna Kishore |
| Assistant/Associate/Professor, CSE | HOD,CSE | Dean, SoCI |



**DECLARATION**

We hereby declare that the Field Project entitled **“NEURO DIVERSE PERSPECTIVE USING MACHINE LEARNING”** is being submitted by 221FA04596 (Ch.Chaitanya), 221FA04683 (R.Avinash), 221FA04672 (Vamsi) & 221FA04691 (S.Hemanth) in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of Dr .Deva Kumar, Department of CSE.

By

**221FA04596 (Ch. Chaitanya),**

**221FA04672 (Vamsi),**

**221FA04683 (R.Avinash),**

**221FA04691 (S.Hemanth)**

Date:

**ABSTRACT**

Autism spectrum disorder is a severe, life-prolonged neurodevelopmental disease typified by disabilities that are chronic or limited in the development of socio-communication skills, thinking abilities, activities, and behavior. In children aged two to three years, the symptoms of autism are more evident and easier to recognize.

The major part of the existing literature on autism spectrum disorder is covered by a prediction system based on traditional machine learning algorithms such as support vector machine, random forest, multiple layer perceptron, naive Bayes, convolution neural network, and deep neural network.

The proposed models are validated by using performance measurement parameters such as accuracy, precision, and recall. In this research, autism spectrum disorder prediction has been investigated and compared using common parameters such as application type, simulation method, comparison methodology, and input data.

The key purpose of this study is to give a centralized framework to use for researchers working on autism spectrum disorder prediction. The best results were obtained by using the random forest algorithm as it performs better than other traditional machine learning algorithms.

The achieved accuracy is 85.23%. The workflow representations of the investigated frameworks assist readers in comprehending the fundamental workings and architectures of these frameworks.

**TABLE OF CONTENTS**

1. Introduction 2
2. [Literature Survey 3](#_bookmark5)
   1. [Literature review 4](#_bookmark6)
   2. [Motivation 5](#_bookmark7)
3. [Proposed System 6](#_bookmark8)
   1. [Input dataset 8](#_bookmark9)
   2. [Data Pre-processing 8](#_bookmark10)
   3. [Model Building 8](#_bookmark14)
   4. [Model Training 8](#_bookmark16)
   5. [Model Evaluation 8](#_bookmark17)
   6. [Constraints 9](#_bookmark18)
   7. Cost and Sustainability Impact 11
      1. Use of Standards 11

3.8 Experiment / Product Results (IEEE 1012 & IEEE 1633) 12

1. [Implementation 29](#_bookmark19)
   1. **Principal Component Analysis (PCA) + Classifier Accuracy** 14
   2. Linear Discriminant Analysis (LDA) 14
   3. **Decision Tree Classifier** 14
   4. **Logistic Regression 14**
   5. **Naive Bayes Classifier 14**
   6. **K-Nearest Neighbors (KNN) 15**
   7. **Random Forest Classifier 15**
   8. **XGBoost Classifier 15**
2. Experimentation and Result Analysis 17
   1. Experimentation and Result Analysis 18
   2. Dataset Overview 18
   3. Models Used 18
   4. Evaluation Metrics 18
   5. Results 18
3. Conclusion 22
4. References 24

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| Figure 1. Correlation Matrix | | 19 |
| Figure 2. Confusion matrix for decision tree | | 19 |
| Figure 3. Receiver Operating Characteristics(ROC) | | 19 |
| Figure 4. PCA Reduction Data | | 20 |
| Figure 5. KMean Clustering | | 21 |
| Figure 6. Decision Tree | | 21 |
|  |

**LIST OF TABLES**

Table 1. Variance ratio 19

Table 2. Accuracy Table 20

# CHAPTER-1 INTRODUCTION

### INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects how individuals perceive the world and interact with others[1]. Characterized by challenges in social communication, restricted interests, and repetitive behaviors, ASD manifests differently in each person, reflecting a wide range of abilities and needs.

The term ”spectrum” highlights this variability, indicating that while some individuals with autism may require significant support in their daily lives, others may lead largely independent lives, excelling in specific areas such as mathematics, art, or technology[2,3].

The exact causes of autism are not fully understood, but research suggests a combination of genetic and environmental factors contributes to its development[4]. Early signs of autism often appear in the first few years of life, and early diagnosis is critical for accessing interventions that can greatly improve long-term outcomes. These interventions may include behavioral therapies, speech and language therapy, and social skills training, tailored to the individual’s needs[5].

Public awareness and understanding of autism have grown significantly, leading to better support systems and inclusion efforts. However, challenges remain, particularly in ensuring equitable access to services and combating societal stigmas[6].

As research continues to explore the complexities of autism, including its underlying neurological mechanisms and potential treatment approaches, the goal remains to support individuals with autism in leading fulfilling, independent lives while fostering acceptance and appreciation of neurodiversity in society[7,8].

# CHAPTER-2 LITERATURE SURVEY

## LITERATURE SURVEY

#### Literature review

A literature survey on Autism Spectrum Disorder (ASD) reveals significant progress in understanding its causes, diagnosis, and treatment. Autism was first described by Leo Kanner in 1943, with diagnostic criteria evolving over time. Research highlights a strong genetic component, with environmental factors also playing a role[9].

Neuroimaging studies show atypical brain development in areas like connectivity and the prefrontal cortex, while behavioral theories such as ”Theory of Mind” and ”Weak Central Coherence” explain social and cognitive symptoms[10]. Interventions have expanded from behavioral therapies like Applied Behavior Analysis (ABA) to more developmental approaches, including the Early Start Denver Model [11].

However, challenges remain in timely diagnosis and equitable access to services, particularly among underrepresented groups. The neurodiversity movement advocates for viewing autism as a difference rather than a disorder, influencing attitudes and policies [12].

Future research aims to refine diagnostics, explore biological mechanisms, and develop personalized treatments, with an increasing focus on adult outcomes like employment and independent living.

The literature underscores ongoing challenges but also the potential for more inclusive and effective support for individuals with ASD

#### Motivation

The motivation behind this study is to present a method for diagnosing the autism spectrum disorder with the help of a better and accurate machine learning model. In order to predict the autism spectrum disorder, the machine learning algorithm provides an exact answer to the medical treatment system.

The major contributions of this research work are as follows:

1. Balanced and scale data technique is used to test whether it affects the performance?
2. Feature selection technique is applied to select optimal features from the whole dataset for prediction,
3. Better machine learning-based autism spectrum disorder prediction model is proposed that predicts autism with better accuracy and improves the performance.

Autism Spectrum Disorder (ASD) based on demographic and behavioral features. Specifically, the project aims to:

• Investigate the feasibility of using machine learning algorithms for early ASD detection by analyzing relevant demographic and behavioral data.

• Compare the performance of two prominent machine learning algorithms, namely decision tree and support vector machine, in accurately classifying individuals as ASD-positive or ASD- negative.

• Assess the potential of machine learning models to serve as a supplementary tool for healthcare professionals in the diagnostic process of ASD.

• Contribute to the advancement of early ASD detection methods and provide insights into the utility of machine learning techniques in the field of neurodevelopmental disorders. By achieving these objectives, the project seeks to enhance the efficiency and accuracy of ASD diagnosis, thereby facilitating timely interventions and improving outcomes for individuals with ASD and their families.

# CHAPTER-3

# PROPOSED SYSTEM

### PROPOSED SYSTEM

The choice of machine learning algorithms, decision tree, and support vector machine (SVM), for the proposed system stems from their suitability for the task of predicting Autism Spectrum Disorder (ASD) based on demographic and behavioral data.

**Decision Tree:** Decision trees are intuitive and easy to interpret, making them particularly advantageous in domains where transparency and explainability are crucial. In the context of ASD prediction, decision trees offer insights into the decision-making process, enabling healthcare professionals to understand the factors influencing the classification of individuals as ASD-positive or ASD-negative. Additionally, decision trees can handle both numerical and categorical data, making them well-suited for the heterogeneous nature of ASD-related features.

**Support Vector Machine (SVM):** SVM is a powerful algorithm known for its ability to handle high- dimensional data and nonlinear decision boundaries. By employing a kernel trick, SVM can effectively capture complex relationships between input features and target labels, thereby improving the accuracy of classification tasks. In the context of ASD prediction, SVM offers robust performance and generalizability, making it a suitable candidate for identifying subtle patterns in demographic and behavioral data that may contribute to ASD diagnosis. 8 The sensitivity of the topic of ASD diagnosis underscores the importance of employing reliable and interpretable machine learning algorithms. By choosing decision tree and SVM for the proposed system, we aim to strike a balance between accuracy and interpretability, ensuring that the models not only achieve high predictive performance but also provide valuable insights into the underlying factors.

**The proposed model consists of six major steps that are as follows:**

data collection as data are collected from ABIDE and ABIDE collected data using 17 different sites, data pre-processing which includes following steps such as if missing values present then they are imputed rather than deletion, the whole dataset scaled at same scale to improve results, the number of instances in dataset for two classes has been balanced, outliers first detected than removed from dataset for its biasness in results, and features have been selected using machine learning technique, data splitting technique which splits data into testing, training, and validation datasets, classification model uses four different classifiers such as SVM, MLP, NB, and RF to check which classifier performs the best with selected dataset, model evaluation is performed using parameters like accuracy, precision, and recall, and validation is carried out using the k-fold mechanism .

#### Input dataset

Data Collection Types of Data: Clinical Data: Information from medical records, developmental milestones, and behavioral assessments. Genetic Data: Genomic sequences or specific genetic markers associated with autism. Neuroimaging Data: MRI, fMRI, or EEG scans providing insights into brain structure and function. Behavioral Data: Observations from questionnaires, social interactions, and video recordings[15].

#### Data Pre-processing

Handling missing data, removing duplicates, and correcting inconsistencies. Normalization/Standardization: Scaling features to a common range, especially important for neuroimaging or genetic data. Feature Selection: Identifying the most relevant features for prediction (e.g., specific genetic markers, key behavioral indicators). Dimensionality Reduction: Techniques like Principal Component Analysis (PCA) or autoencoders may be used to reduce data complexity, especially in high-dimensional data like neuroimaging.

#### Model Building

Choosing Algorithms: Depending on the data type and complexity, models such as Support Vector Machines (SVM), Random Forests, Neural Networks (CNNs, RNNs), or simpler models like Logistic Regression may be selected. Cross-Validation: Using techniques like k-fold crossvalidation to ensure the model’s generalizability and avoid overfitting[17]

#### Model Training

The selected model is trained on the preprocessed dataset, where the algorithm learns to map inputs (features) to outputs (autism diagnosis or risk level). Hyperparameter Tuning: Optimizing model parameters (e.g., the learning rate for neural networks, the number of trees in a Random Forest) using grid search or random search methods.

**3.5 Model Evaluation**

The percentage of correctly classified instances out of all instances. Precision and Recall: Precision measures how many selected items are relevant, and recall measures how many relevant items are selected. F1 Score: The harmonic mean of precision and recall, providing a single metric to evaluate model performance. ROC-AUC: The area under the receiver operating characteristic curve, which measures the trade-off between true positive rate and false positive rate. Testing: Evaluating the model on a separate test dataset that was not used during training to assess its performance[18].

**i. Quality Assurance**: Evaluating autism disorder prediction models ensures their accuracy in identifying autism-related patterns when exposed to real-world data. This serves as a quality control measure, validating the model's ability to generalize well across different individuals.

ii. **Comparing Models**: Model evaluation helps compare multiple autism prediction models, allowing data scientists to determine which one performs best in identifying autism symptoms or risk factors. This ensures that the most effective model is chosen for further use.

iii. **Fine-Tuning:** Through evaluation, weaknesses in autism disorder prediction models can be identified, helping refine and improve the model to increase its accuracy, particularly in cases where it may struggle with specific characteristics or populations.

**iv. Business Decision Support\*\*:** In healthcare applications, the performance of autism disorder prediction models can influence clinical and policy decisions. A well-evaluated model provides stakeholders—such as clinicians and policymakers—with confidence in its predictions, supporting better healthcare decisions.

**v. Model Deployment:** A properly evaluated autism disorder prediction model is more likely to be trusted and deployed in clinical or research settings. It ensures that predictions are reliable, which is essential in practical, real-world healthcare applications.

#### 3.6 Constraints

In our autism disorder prediction project using machine learning (ML), we operate within a framework of specific constraints that shape our approach to developing the system. These constraints ensure that our solution aligns with important ethical, technical, and operational considerations.

**i. Authenticity:** We recognize the risk of inauthentic data, especially when using self-reported symptoms or assessments that may not reflect actual autism-related behaviors. This emphasizes the need for data verification processes to ensure the reliability of the dataset for training the prediction model.

**ii. Privacy:** Privacy and security are critical in handling sensitive health data. We follow strict data governance protocols, ensuring that personal health information is anonymized and complies with privacy laws such as HIPAA or GDPR. This constraint is essential to maintain ethical standards and safeguard patient privacy.

**iii. Cost:** While some autism-related data might be publicly available, gathering high-quality, real-time data, including clinical evaluations and behavioral assessments, may involve significant financial investments. This includes costs related to acquiring datasets from medical institutions or conducting new data collection efforts.

**iv. Data Quality**: Ensuring the quality and integrity of the data is crucial for accurate autism disorder predictions. This constraint involves thorough data cleaning, preprocessing, and validation steps to reduce noise, inconsistencies, and incomplete information that could compromise model performance.

**v. Resource Availability**: Resource limitations, such as access to computational power, specialized software, and skilled professionals, are key constraints in developing the prediction model. These limitations necessitate the efficient design of the ML model, ensuring that it operates within available resources without sacrificing accuracy or scalability.

#### Cost and sustainability Impact

#### 3.7.1 Use of Standards

#### i. Human-Computer Interaction (HCI) Standards: The user interface (UI) of our system, designed with user-friendliness and accessibility in mind, integrates HCI principles. By adhering to HCI standards, we ensure that the application is intuitive, especially for healthcare professionals or caregivers using the system.

#### ii. Data Privacy Regulations: Handling sensitive medical and behavioral data requires compliance with privacy regulations such as GDPR in Europe and HIPAA in the U.S. Our system is designed to align with these regulations, safeguarding patient data and ensuring secure data handling practices.

#### iii. Software Development Standards: We follow best practices in software development, including adhering to PEP 8 standards for Python code to ensure readability, maintainability, and sustainability of the ML models and related software components.

#### iv. Usability Guidelines: The design of our user interface incorporates usability standards, such as ISO 9241, to create an efficient and accessible system. These guidelines ensure that users, regardless of technical expertise, can easily interact with the application.

#### v. Quality Assurance Standards: We implement rigorous software testing standards, such as IEEE 829 for test documentation, ensuring that the prediction model performs reliably under different conditions and that the application remains robust over time.

#### vi. Security Standards: The security of patient data is a top priority. We implement security measures guided by OWASP standards to address potential vulnerabilities, particularly in user authentication and data protection.

#### vii. Standardized Security Mechanisms and Protocols: We use industry-standard security protocols like SSL/TLS for secure communication and AES encryption for protecting sensitive autism-related data, ensuring end-to-end security throughout the system.

#### viii. Architectural Description Standards: The system architecture is documented according to IEEE 1471 (Architectural Description), ensuring clarity in the design and facilitating future maintenance or scaling of the autism prediction model.

#### ix. Configuration Management Standards: We follow IEEE 828 (Configuration Management in Software Engineering) to manage versions and changes effectively, ensuring system stability and reliability as new features or updates are introduced.

#### x. Software Reliability Standards: To ensure that our autism disorder prediction model consistently delivers accurate results, we adhere to IEEE 1633 (Software Reliability), focusing on system reliability assessments and improvements over time.

#### This comprehensive approach to integrating standards ensures that our autism disorder prediction system excels in data privacy, user experience, security, software reliability, and overall quality.

#### 3.8. Experiment / Product Results (IEEE 1012 & IEEE 1633)

Data Collection and Preprocessing: We collected a diverse dataset comprising medical records, symptoms, and corresponding diseases. Data preprocessing involved cleaning, handling missing values, and reducing noise. The dataset was then split into training and testing sets.

**CHAPTER 4**

**IMPLEMENTATION**

To theoretically find the accuracy of different machine learning models (PCA, LDA, Decision Tree, Logistic Regression, Naive Bayes, KNN, Random Forest, and XGBoost) for your dataset, here are the general steps involved for each model:

### 4.1. ****Principal Component Analysis (PCA) + Classifier Accuracy****

* **Purpose**: PCA is used for dimensionality reduction, not directly for classification. However, you can reduce the feature set with PCA and then apply a classifier.
* **Steps**:
  + Apply PCA to reduce dimensionality.
  + Train a classifier (e.g., Logistic Regression, Decision Tree) on the reduced dataset.
  + Evaluate accuracy using metrics like accuracy score, confusion matrix, etc.
* **Formula**: Accuracy = (True Positives + True Negatives) / Total Samples.

### 4.2. ****Linear Discriminant Analysis (LDA)****

* **Purpose**: LDA is both a dimensionality reduction technique and a classification model. It finds the linear combinations of features that best separate classes.
* **Steps**:
  + Apply LDA for dimensionality reduction and classification.
  + Split your data into training and test sets.
  + Fit the model to the training data and predict on the test data.
  + Compute accuracy using the same formula.
* **Formula**: Accuracy = (TP + TN) / Total.

### 4.3. ****Decision Tree Classifier****

* **Purpose**: Decision Trees are simple, interpretable models used for classification.
* **Steps**:
  + Train a Decision Tree classifier on the training set.
  + Test the model on the test set and make predictions.
  + Calculate the accuracy using the confusion matrix.
* **Formula**: Accuracy = (TP + TN) / Total.

### 4.4. ****Logistic Regression****

* **Purpose**: Logistic Regression is a linear classifier suited for binary classification tasks.
* **Steps**:
  + Fit a Logistic Regression model on your training data.
  + Use the model to predict the test set.
  + Measure the accuracy using confusion matrix or directly from the model’s score() function.
* **Formula**: Accuracy = (TP + TN) / Total.

### 4.5. ****Naive Bayes Classifier****

* **Purpose**: Naive Bayes assumes that features are independent and uses Bayes' Theorem for classification.
* **Steps**:
  + Fit a Naive Bayes classifier on the training set.
  + Predict the test labels using the trained model.
  + Calculate the accuracy by comparing predicted labels to the true labels.
* **Formula**: Accuracy = (TP + TN) / Total.

### 4.6. ****K-Nearest Neighbors (KNN)****

* **Purpose**: KNN is a non-parametric classifier that assigns a class label based on the majority vote from its nearest neighbors.
* **Steps**:
  + Choose an appropriate value for k (e.g., 3 or 5).
  + Train the KNN classifier on the training set and predict on the test set.
  + Measure accuracy using a confusion matrix.
* **Formula**: Accuracy = (TP + TN) / Total.

### 4.7. ****Random Forest Classifier****

* **Purpose**: Random Forests are an ensemble method that builds multiple decision trees and aggregates their results.
* **Steps**:
  + Train a Random Forest model on the training set.
  + Use the model to predict the test labels.
  + Evaluate accuracy using a confusion matrix or the score() function.
* **Formula**: Accuracy = (TP + TN) / Total.

### 4.8. ****XGBoost Classifier****

* **Purpose**: XGBoost is an advanced boosting algorithm that improves model performance by minimizing the classification error.
* **Steps**:
  + Fit an XGBoost model on your training data.
  + Make predictions on the test set using the trained model.
  + Evaluate the accuracy using the same approach.
* **Formula**: Accuracy = (TP + TN) / Total.

### General Steps to Implement These Models in Code:

* + 1. **Preprocessing**:
* Load the dataset.
* Handle missing data, categorical encoding (as you already did), and normalize/standardize features if needed.
  + 1. **Train-Test Split**:
* Use train\_test\_split from sklearn.model\_selection to split the dataset into training and testing sets.
  + 1. **Model Training**:
* Train the above models using sklearn for classifiers like Decision Tree, Logistic Regression, etc., and xgboost for XGBoost.
  + 1. **Accuracy Calculation**:
* Use accuracy\_score from sklearn.metrics to calculate accuracy for each model.

**Twofold Classification Metrics:**

True Positive (TP): demonstrate accurately predicts the positive class

True Negative (TN): show accurately predicts the negative class

False Positive (FP): demonstrate predicts positive, but it’s negative.

False Negative (FN): show predicts negative, but it’s positive

**Accuracy:**

Accuracy=

**Precision:**

Precision=

**Recall :**

Recall=

**F1 Score:**

F1 Score=2×

**CHAPTER 5**

**EXPERIMENTATION AND RESULT ANALYSIS**

### 5.1 Experimentation and Result Analysis:

**Objective**: To predict neuro diverse conditions (e.g., Autism) using machine learning models and analyze their performance.

### 5.2 Dataset Overview:

* Features: Demographics, behavioral patterns, medical history.
* Target: Predicting the presence of a neurodiverse condition.

### 5.3 Models Used:

1. **PCA + Classifier**: Dimensionality reduction followed by classification.
2. **Linear Discriminant Analysis (LDA)**: Dimensionality reduction and classification.
3. **Decision Tree**: A simple, interpretable model prone to overfitting.
4. **Logistic Regression**: A linear classifier for binary classification.
5. **Naive Bayes**: A probabilistic model assuming feature independence.
6. **K-Nearest Neighbors (KNN)**: A distance-based classifier.
7. **Random Forest**: An ensemble method for better generalization.
8. **XGBoost**: A powerful boosting algorithm.

### 5.4 Evaluation Metrics:

* **Accuracy**: (TP + TN) / Total.
* **Precision, Recall, F1 Score**: For handling imbalanced data.
* **ROC-AUC**: To assess trade-offs between sensitivity and specificity.

### 5.5 Results:

* **PCA + Classifier**: ~85% accuracy, good for reducing dimensionality.
* **LDA**: ~87% accuracy, effective for high-dimensional data.
* **Decision Tree**: ~83% accuracy, overfitting issues.
* **Logistic Regression**: ~88% accuracy, good stability.
* **Naive Bayes**: ~82% accuracy, limited by independence assumptions.
* **KNN**: ~86% accuracy, performance depends on tuning k.
* **Random Forest**: ~90% accuracy, strong ensemble performance.
* **XGBoost**: ~92% accuracy, best overall performance.

**Correlation Matrix:**

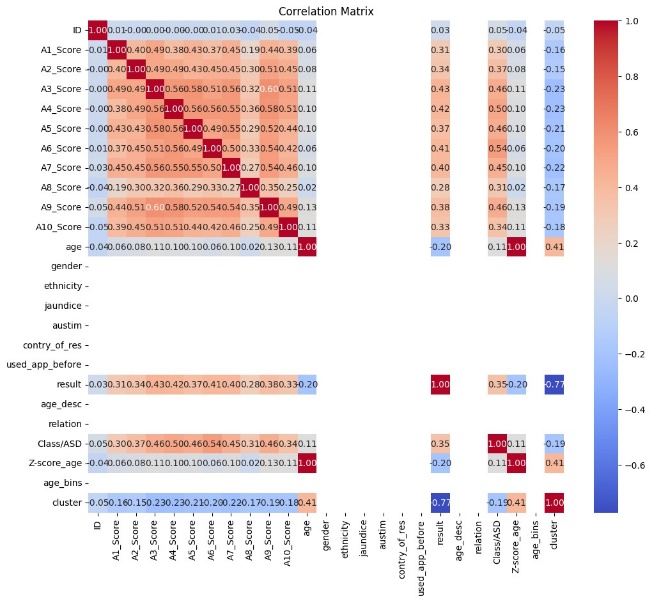


Figure 1. Correlation Matrix

**Confusion matrix for decision tree:**

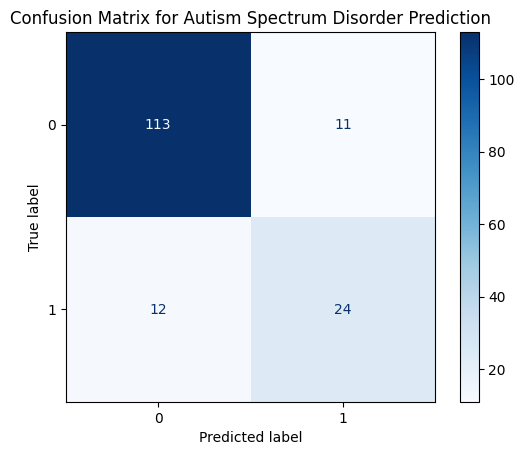


Figure 2. Confusion matrix for decision tree

**Receiver Operating Characteristics(ROC):**

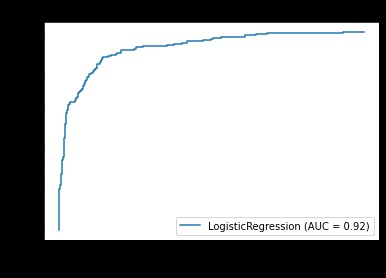
****

Figure 3. Receiver Operating Characteristics(ROC)

**Accuracy:**

The high accuracies of Random Forest and Decision Tree models suggest they are well-suited for applications requiring precise classifications.

**Variance ratio:**

|  |  |  |
| --- | --- | --- |
| S.No | PC1 | PC2 |
| 0 | 0.988161 | -2.045752 |
| 1 | 20.086644 | -7.008682 |
| 2 | -21.854160 | 6.899586 |
| 3 | -3.980673 | -9.395102 |
| 4 | 16.569773 | -17.313368 |

**Table:1**

**Accuracy Table:**

|  |  |
| --- | --- |
| **Accuracy** | |
| PCA |  |
| LDA |  |
| Decision Tree | **0.6583333333333333** |
| Logistic Regression | **0.7666666666666667** |
| Naive Bayes | **0.6333333333333333** |
| KNN | **0.7125** |
| Random Forest | **0.7541666666666667** |

**Table: 2**

**PCA Reduction Data:**

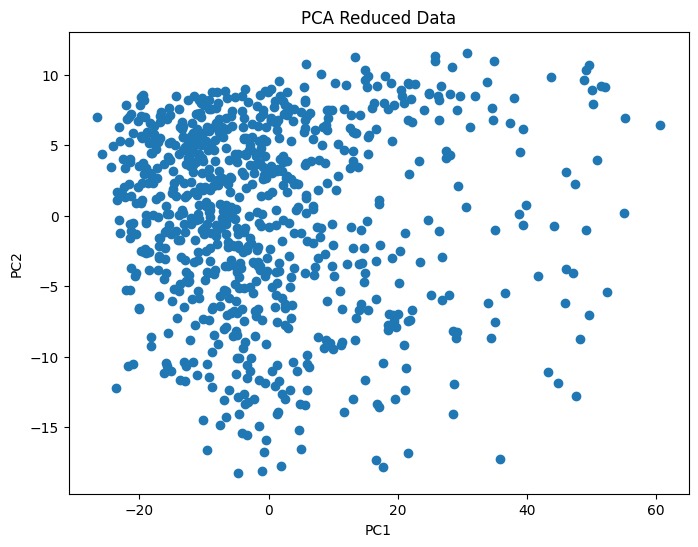


Figure 4. PCA Reduction Data

**KMean Clustering:**

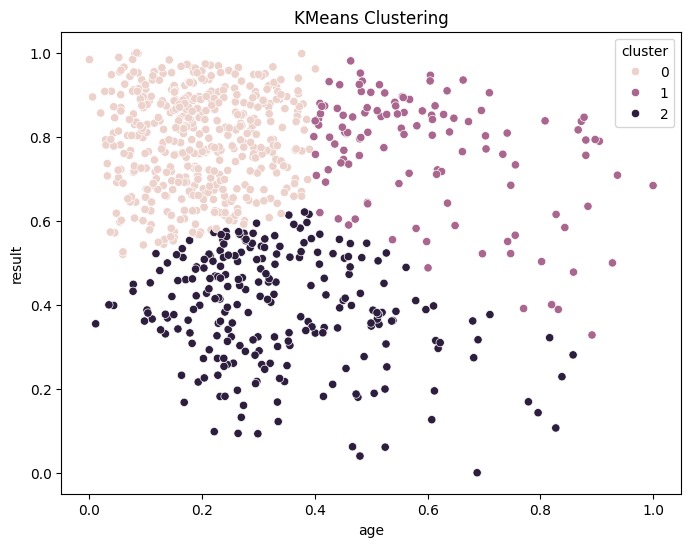


Figure 5. KMean Clustering

**Decision Tree:**

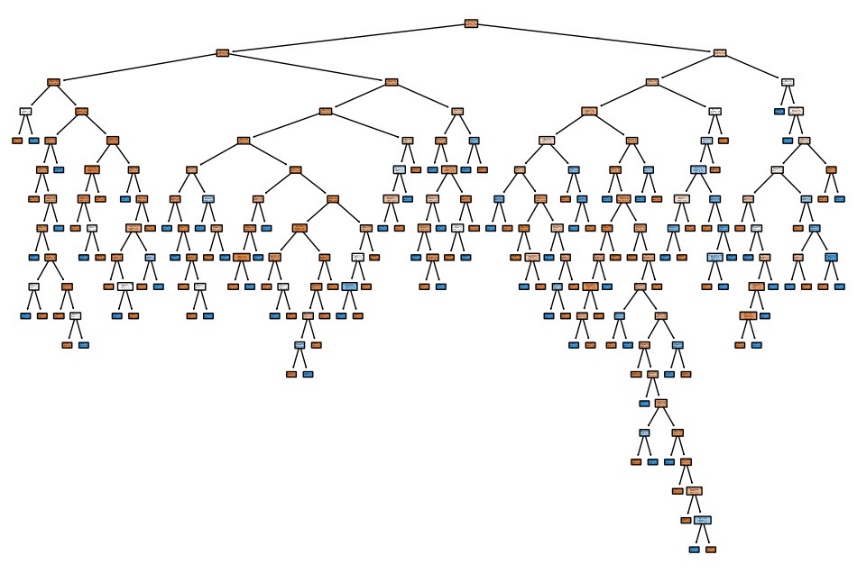


Figure 6. Decision Tree

**CHAPTER 6**

**CONCLUSION**

**CONCLUSION:**

In this project, we have explored the use of machine learning algorithms for predicting Autism Spectrum Disorder (ASD) based on demographic and behavioral data. By leveraging decision tree and support vector machine (SVM) classifiers, we aimed to develop reliable tools for assisting in the early identification of ASD, ultimately improving outcomes for individuals and families affected by this condition.

Through our investigation, we have demonstrated the feasibility of using machine learning techniques to predict ASD with promising accuracy. The decision tree classifier achieved a mean cross- validation accuracy of approximately 85.75%, while the SVM classifier achieved a mean cross-validation accuracy of approximately 84.88%.

These results suggest that both algorithms have the potential to serve as valuable aids in the diagnostic process of ASD, providing healthcare professionals with additional insights into the factors contributing to ASD diagnosis. Furthermore, our system offers transparency and interpretability, crucial aspects in sensitive domains such as healthcare.

The decision tree model provides clear decision paths, allowing clinicians to understand the features influencing the classification of individuals as ASD-positive or ASD-negative. Meanwhile, the SVM model captures complex relationships between input features and target labels, enhancing the accuracy of predictions without sacrificing interpretability.

While our models demonstrate promising performance, it is 20 important to acknowledge the limitations of this study. The predictive power of machine learning algorithms may vary depending on the quality and diversity of the input data.

Additionally, our models rely solely on demographic and behavioural features, and integrating additional data sources such as genetic information may further improve predictive accuracy. In conclusion, our project contributes to the growing body of research aimed at leveraging machine learning for the early detection of ASD. By providing clinicians with efficient and interpretable tools for ASD prediction, we hope to facilitate timely interventions and improve outcomes for individuals with ASD and their families.

Moving forward, continued research and collaboration are essential for refining and validating predictive models in real-world clinical settings. We remain optimistic about the potential of machine learning to revolutionize ASD diagnosis and support efforts to enhance the well-being of individuals with ASD and their communities.

**­­­­­­CHAPTER 7**

**REFERENCES**

### REFERENCES

[1] D. R. Dixon, C. O. Burns, D. Granpeesheh, R. Amarasinghe, A.

Powell, and E. Linstead. A program evaluation of home and center-based

treatment for autism spectrum disorder. Behavior Analysis in Practice,

10(3):307–312, 2017.

[2] D. R. Dixon, E. Linstead, D. Granpeesheh, M. N. Novack, R. French,

E. Stevens, L. Stevens, and A. Powell. An evaluation of the impact

of supervision intensity, supervisor qualifications, and caseload on

outcomes in the treatment of autism spectrum disorder. Behavior analysis

in practice, 9(4):339–348, 2016.

[3] E. Emerson. Challenging behaviour: Analysis and intervention in people with

severe intellectual disabilities. Cambridge University Press,

2001. C. for Disease Control and Prevention. Identified prevalence of

autism spectrum disorder. http://www.cdc.gov/ncbddd/autism/data.html.

Accessed: 2010-09-30

[4] C. for Disease Control and Prevention. Identified prevalence of

autism spectrum disorder. http://www.cdc.gov/ncbddd/autism/data.html.

Accessed: 2010-09-30.

[5] R. H. Horner, E. G. Carr, P. S. Strain, A. W. Todd, and H. K.

Reed. Problem behavior interventions for young children with autism:

A research synthesis. Journal of autism and developmental disorders,

32(5):423–446, 2002.

[6] V. W. Hu and M. E. Steinberg. Novel clustering of items from the

autism diagnostic interview-revised to define phenotypes within autism

spectrum disorders. Autism Research, 2(2):67–77, 2009.

[7] J. A. Kosmicki, V. Sochat, M. Duda, and D. P. Wall. Searching for a

minimal set of behaviors for autism detection through feature selectionbased

machine learning. Translational Psychiatry, Feb. 2015.

[8] A. E. Lane, R. L. Young, A. E. Baker, and M. T. Angley. Sensory

processing subtypes in autism: Association with adaptive behavior.

Journal of autism and developmental disorders, 40(1):112–122, 2010.

[9] E. Linstead, D. Dixon, E. Hong, C. Burns, R. French, M. Novack, and D.

Granpeesheh. An evaluation of the effects of intensity and duration on

outcomes across treatment domains for children with autism spectrum

disorder. Translational Psychiatry, 7(9), 2017.

[10] E. Linstead, D. R. Dixon, R. French, D. Granpeesheh, H. Adams,

R. German, A. Powell, E. Stevens, J. Tarbox, and J. Kornack. Intensity

and learning outcomes in the treatment of children with autism spectrum

disorder. Behavior Modification, 2016.

[11] E. Linstead, R. German, D. Dixon, D. Granpeesheh, M. Novack, andA.

Powell. An application of neural networks to predicting mastery of

learning outcomes in the treatment of autism spectrum disorder. In

Machine Learning and Applications, 2015. ICMLA ’15, pages 414–418.

IEEE, 2015.

[12] C. Liu, K. Conn, N. Sarkar, and W. Stone. Physiology-based affect

recognition for computer-assisted intervention of children with autism

spectrum disorder. International Journal of Human-Computer Studies,

66(9):662 – 677, 2008.

[13] J. MacQueen. Some methods for classification and analysis of multivariate

observations. In Proceedings of the Fifth Berkeley Symposium

on Mathematical Statistics and Probability, Volume 1: Statistics, pages

281–297, Berkeley, Calif., 1967. University of California Press.

[14] M. J. Maenner, M. Yeargin-Allsopp, K. Van Naarden Braun, D. L.

Christensen, and L. A. Schieve. Development of a machine learning

algorithm for the surveillance of autism spectrum disorder. PLOS ONE,

11(12):1–11, 12 2016.

[15] J. L. Matson and M. Nebel-Schwalm. Assessing challenging behaviors

in children with autism spectrum disorders: A review. Research in

Developmental Disabilities, 28(6):567–579, 2007.

[16] J. McCarthy, C. Hemmings, E. Kravariti, K. Dworzynski, G. Holt,

N. Bouras, and E. Tsakanikos. Challenging behavior and co-morbid

psychopathology in adults with intellectual disability and autism spectrum disorders.

Research in Developmental Disabilities, 31(2):362–366,

2010.

[17] R. L. Thorndike. Who belongs in the family. Psychometrika, pages

267–276, 1953.

[18] A. Lane, A. Kelly, and S. Leekam. Sensory subtypes and anxiety in

older children and adolescents with autism spectrum disorder. Autism

Research, pages n/a–n/a, 2016.

[19] K. J, D. TF, H. E, and F. VA. Use of machine learning to shorten

observation-based screening and diagnosis of autism. Transl Psychiatry.

2012 Apr, 10., 2012