AUTISM PREDICTION MODEL IN CHILDREN

Enhancing Early Detection and Intervention

AGENDA

Overview

Business Understanding

Data understanding and Data preparation

Modeling

Evaluation

Recommendations and Next Steps

OVERVIEW

This project aims to predict Autism Spectrum Disorder (ASD) in children by analyzing various features such as demographic information, medical history, and screening scores. The model will evaluate factors like age, gender, ethnicity, family history of ASD, and results from behavioral questionnaires to assess the likelihood of a child having ASD. By integrating these diverse data points, the model seeks to enhance early detection and support timely interventions for better long-term outcomes.

BUSINESS UNDERSTANDING

Business Problem

Autism Spectrum Disorder is often underdiagnosed, especially in younger children or those from diverse backgrounds. Early diagnosis can significantly improve intervention outcomes. **Springwell Diagnostics** wants a predictive model that identifies the likelihood of ASD early on, so it can be incorporated into their products and services, increasing market penetration and generating revenue.

BUSINESS UNDERSTANDING

Stakeholder Audience: Springwell Diagnostics, health providers, parents, and public health organizations.

Objectives

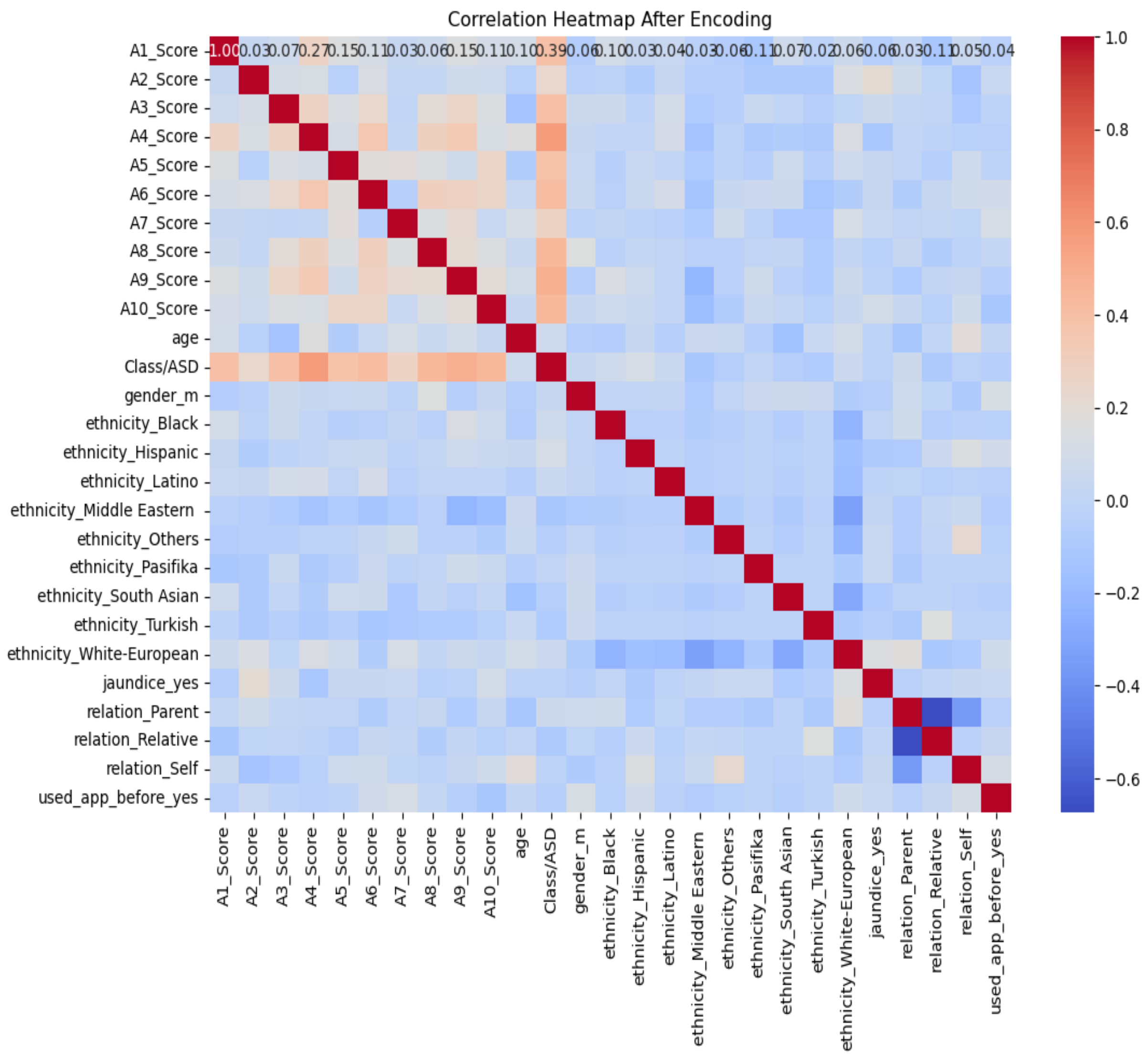
1) Evaluate ROI for Springwell Diagnostics.
2) Enhance screening for healthcare providers.
3) Offer accessible screening for parents and educators.
4) Support public health organizations and insurers.

DATA UNDERSTANDING

Dataset: 292 records of autism screening results for children aged 4-11 years.

Key Features:

- Demographics: Age, gender, ethnicity.
- Medical History: Family ASD history, jaundice.
- Screening Results: Scores from behavioral questions.



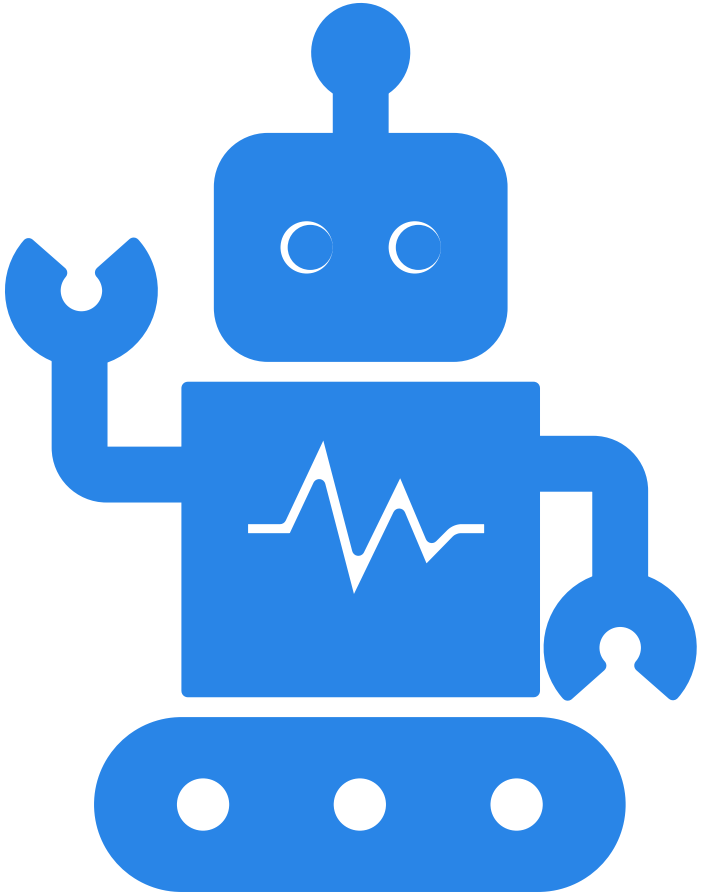
CORRELATION HEAT MAP

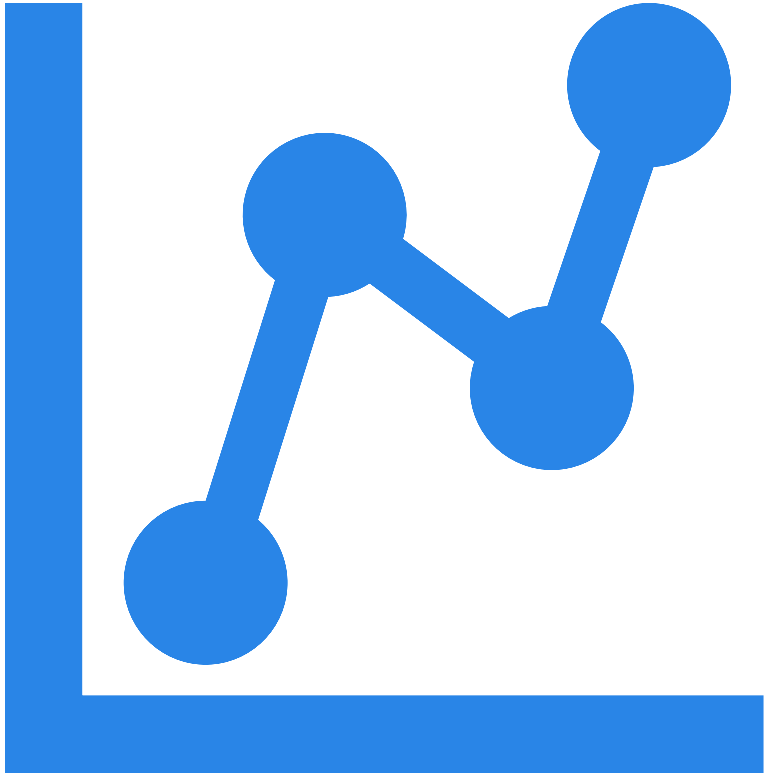
Interpretation:

- The A_scores show strong correlation and thus play a key role in establishing a robust baseline model. This helps in setting a solid foundation for evaluating and refining the model as additional complexity is introduced.

MACHINE LEARNING

Machine learning is a branch of artificial intelligence that enables systems to learn from data and improve their performance over time without being explicitly programmed. It involves training algorithms to recognize patterns and make predictions or decisions based on new, unseen data.





MODELING

Three machine learning models were utilized:

- Logistic Regression: Simple, effective classification.
- Decision Tree: Handles non-linear relationships.
- Random Forest: Aggregates multiple trees to improve performance.

Baseline Model Choice:

Target Variable: Class/ASD

Independent Variable: A10_Score

MODEL PERFORMANCE

Metrics	Logistic Regression	Decision Tree	Random Forest
Accuracy	93.2%	86.4%	91.5%
Confusion Matrix	36 True Negatives 3 False Positives 1 False Negative 19 True Positives	33 True Negatives 6 False Positives 2 False Negatives 18 True Positives	36 True Negatives 3 False Positives 2 False Negatives 18 True Positives
Classification Report	High precision, recall, and F1-score for positive class, indicating strong performance in identifying autism cases.	Lower precision and recall for both classes compared to Logistic Regression, indicating less reliable performance.	Comparable precision, recall, and F1-score to Logistic Regression, showing effective balance between bias and variance.

EVALUATION

Predictive Model Chosen

Logistic Regression is chosen due to its highest accuracy (93.2%) and strong performance in identifying autism cases, as indicated by its high precision, recall, and F1-score for the positive class. This model outperforms the Decision Tree and Random Forest in accuracy and provides more reliable predictions, making it the most effective choice for the task.

ENSURING MODEL ROBUSTNESS

1. **Cross-Validation:**

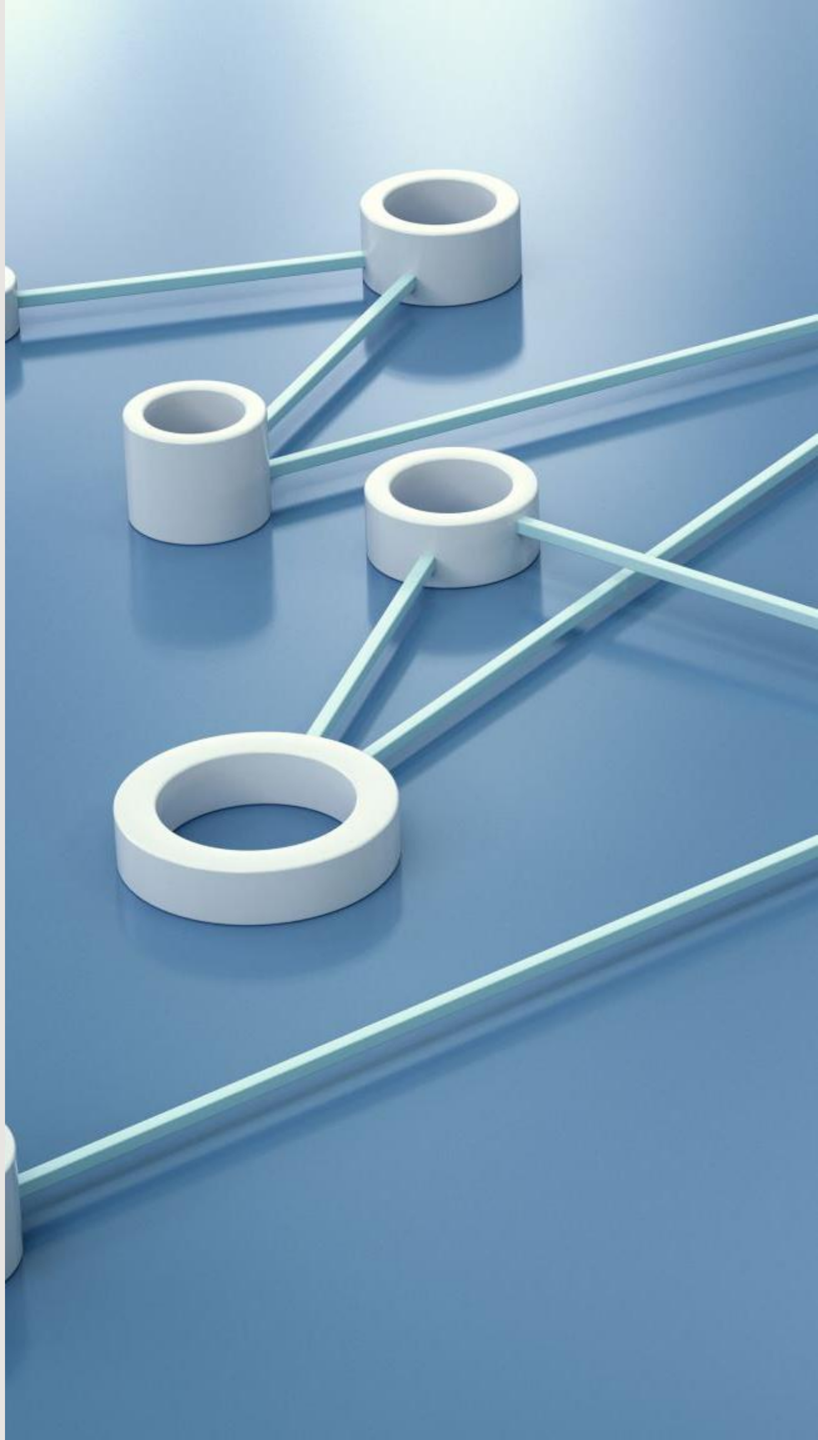
   We used cross-validation to evaluate the model on multiple data subsets, ensuring robust performance and generalization to new data.

2. **Learning Curves:**

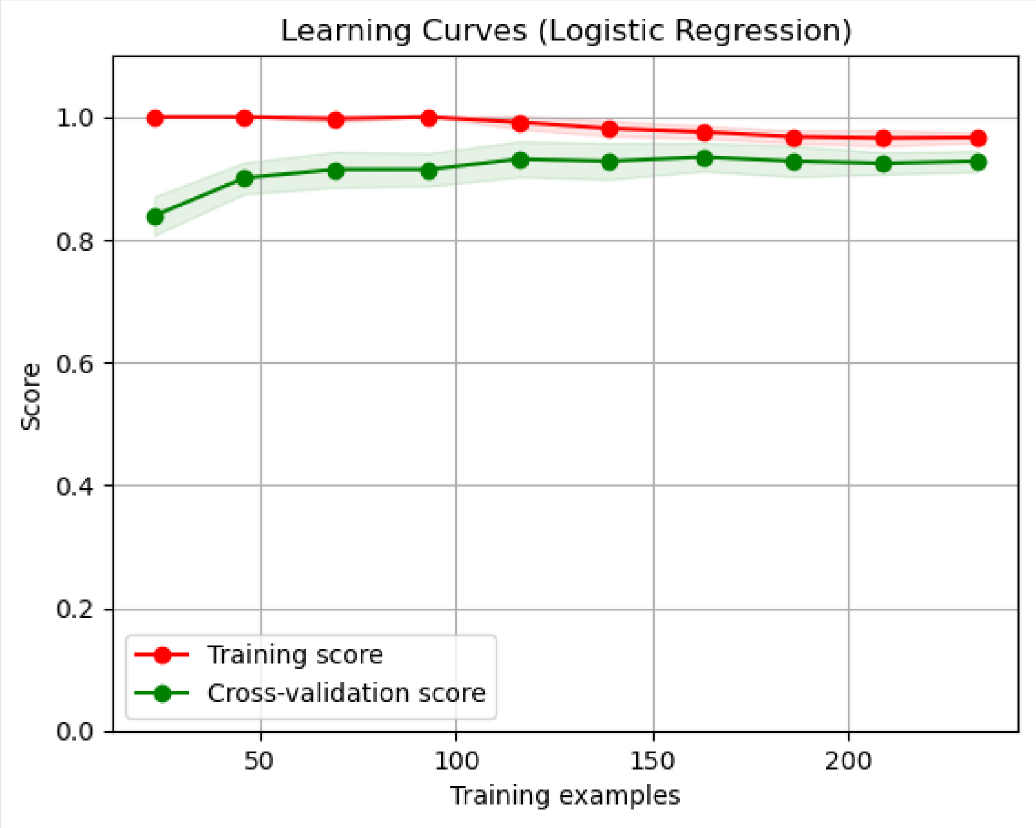
   Analyzed learning curves to track training and validation scores, identifying and addressing potential overfitting or underfitting issues.

3. **Model Complexity:**

   Compared different models to balance complexity and performance, avoiding models that were too simple or too complex.



LEARNING CURVE



- ***Interpretation:*** The initial low validation score compared to the high training score suggested potential overfitting. However, as the dataset grew, the validation score improved, reflecting better generalization and reduced overfitting.

RECOMMENDATIONS

Feature Engineering: Explore additional features and interactions so as to improve the model

ROI Evaluation: Assess market demand and potential revenue from tool licensing.

Tool Integration: Develop user-friendly applications for healthcare providers and parents.

Public Health Support: Create bulk licensing options for community-wide screenings.

NEXT STEPS

FURTHER RESEARCH: EXPLORE ADVANCED MODELS AND ADDITIONAL DATA SOURCES.

PARTNERSHIPS: ENGAGE WITH HEALTHCARE PROVIDERS FOR REAL-WORLD TESTING.

DEPLOYMENT: DEVELOP A PROTOTYPE AND PILOT THE TOOL IN SELECTED CLINICS.

THANK YOU FOR YOUR TIME

- **ANY QUESTIONS?**

- **Elsie Serem**
- **LinkedIn: www.linkedin.com/in/elsie-serem-237b26287**