

**DSCI/STAT 8750: Data Science Capstone Project Proposal**

---

<b>Project Title</b>	Modeling Early Autism Detection in Children
<b>Group</b>	B
<b>Student Names</b>	1. Sandra Lopez Padilla 2. Kalyan Ghimire 3. Clion Muhoza 4. Olivier Niyonshuti Mizero

**Background:** *Describe the problem or topic area the project will address*

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects communication, social interactions, and behavior in children. ASD affects about 1% of the population (Vakadkar et al., 2021) and is often not diagnosed in a timely or efficient manner. Early diagnosis is crucial because timely intervention can significantly improve developmental outcomes (Hyman et al., 2020). However, diagnosing ASD at an early stage remains a challenge due to the wide range of symptom variations and the reliance on expert assessment. Barbaro and Halder (2023) emphasize that these challenges are further compounded by the lack of standardized screening tools and limited awareness among both caregivers and healthcare professionals, making early detection even more difficult.

Our project aims to address some of these challenges using data-driven methods. We will analyze the “Autistic Spectrum Disorder Screening Data for Children” dataset from the UCI Machine Learning Repository, which includes responses to screening questions along with demographic information. By applying predictive modeling, we seek to identify factors that contribute to early autism detection and improve the accuracy of classification models used for ASD screening.

**Significance:** *Explain why this problem is important and relevant*

Early identification of autism can help parents and healthcare providers take timely actions to support a child’s development. Machine learning models have the potential to enhance traditional screening methods by making predictions based on patterns found in real-world data (Ayub et al., 2024). Our study will contribute to the field by exploring advanced modeling techniques, improving detection accuracy, and providing insights that could aid in developing more effective screening tools.

**Objective:** *Define the specific goals or questions the project aims to address*

Our team's goal is to develop a predictive model with higher accuracy for early autism detection using machine learning techniques. Specifically, we aim to:

- Perform exploratory analysis (EDA) to understand the dataset, identify missing values, and detect potential biases.
- Implement data preprocessing techniques to improve data quality and feature engineering to get the most out of our data.
- Conduct literature review to examine existing research on autism prediction models and identify commonly used algorithms.
- Explore and compare different machine learning models to improve detection accuracy.
- Explore clustering techniques to uncover hidden patterns in the dataset that may enhance predictive performance.
- Evaluate our final model's performance using accuracy, precision, recall, and F-1 score.
- Identify which attributes or features are the most important for predicting early detection of autism according to our best and final model.
- Provide conclusions and recommendations on how our findings can contribute to more effective early autism detection.

### References

1. Vakadkar, K., Purkayastha, D. & Krishnan, D. Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques. *SN COMPUT. SCI.* 2, 386 (2021). <https://doi.org/10.1007/s42979-021-00776-5>
2. Hyman, S. L., Levy, S. E., Myers, S. M., Kuo, D. Z., Apkon, S., Davidson, L. F., ... & Bridgemohan, C. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics*, 145(1).
3. Barbaro, J., & Halder, S. (2016). Early identification of autism spectrum disorder: Current challenges and future global directions. *Current Developmental Disorders Reports*, 3, 67-74.
4. Ayub, R., Ansarullah, S. I., Samdani, F., Akhtar, A., Afzal, M., & Ahmed, S. M. (2024). Enhancing early detection of autistic spectrum disorder in children using machine learning approaches. *Journal of King Saud University-Science*, 36(10), 103468.