

Logistic Regression - Student Learning Guide

Autism Prediction ML Project Presentation Script

Hello everyone,

Today, I'll walk you through a machine learning project focused on predicting autism spectrum disorder (ASD) using screening questionnaire data.

This project demonstrates the full ML lifecycle - from data exploration and cleaning to model building and evaluation.

1. Project Objective

Project Objective:

Our goal is to build a classification model that identifies individuals who may be on the autism spectrum, using behavioral responses and personal background information.

This is especially useful for early screening and support planning.

2. Dataset Overview

Dataset Overview:

- Data includes 10 behavioral questions (A1-A10), medical indicators (jaundice, autism history), age, country, and gender.
- The target variable is `Class/ASD`, where 1 indicates likely ASD and 0 indicates non-ASD.

3. Data Cleaning & Preparation

Data Cleaning & Preparation:

- Missing values and non-numeric types were handled.
- Feature engineering created new columns like `sum_score` and `ind` to better capture behavior patterns.
- We used log transformation on age and created age groups for interpretability.
- Categorical features were encoded with one-hot encoding.

4. Modeling

Logistic Regression - Student Learning Guide

Modeling:

We trained three models: Logistic Regression, Support Vector Machine (SVM), and XGBoost.

- Logistic Regression served as a strong baseline.
- SVM achieved very high recall, reducing the chance of missing true ASD cases.
- XGBoost gave the best overall performance with perfect recall and high precision.

5. Evaluation

Evaluation Metrics:

- We used accuracy, precision, recall, and F1 score.
- Recall was prioritized to ensure actual ASD cases were not missed.
- XGBoost achieved 95% accuracy and 100% recall, making it the best fit for this use case.

6. Key Takeaways

Key Takeaways:

- Feature engineering and data cleaning are just as important as the model choice.
- Logistic regression is interpretable and fast - a great starting point.
- Evaluation metrics should match your goal - in healthcare, missing a case (low recall) is worse than a false positive.
- Tree-based models like XGBoost often outperform others when tuned well.

7. Closing

Thank you for your time!

This project demonstrates how machine learning can be applied to sensitive areas like autism screening, and the importance of ethical, accurate modeling.

Feel free to explore the notebook and learning guides to deepen your understanding.