# Capstone Project Summary: Predicting Academic Achievement Using Cognitive and Social-Emotional Data

Author:  
Evelyn Johnson

Jupyter Notebook Link:  
<https://github.com/eviesjohnson/Capstone-3>

## Project Overview

This project explores how cognitive reasoning ability (measured by the CogAT) and social-emotional competence (measured by the DESSA) predict student academic achievement on state assessments in reading and math. The goal was to evaluate whether machine learning models can improve our ability to identify students who may need additional instructional or social-emotional supports.

## Methods and Modeling

The dataset included student-level data from approximately 2,300 students, including CogAT subtest scores, DESSA scale scores, and state assessment results in reading and math. Student demographics (e.g., gender, EL, SPED, Gifted, and race) were also included.  
  
I trained and evaluated several supervised learning models:  
- Multiple Linear Regression  
- LASSO and Ridge Regression  
- Random Forest Regressor  
  
Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R² Score. All models were trained on 80% of the dataset and tested on the remaining 20%.

## Key Findings

- Linear regression outperformed Random Forest in both reading and math prediction tasks, suggesting that relationships among predictors and outcomes are largely linear.  
- CogAT subtests (especially Verbal for reading and Quantitative for math) were among the strongest predictors of achievement.  
- DESSA scores played a meaningful role, especially when exploring performance beyond cognitive ability.  
  
To better understand this, I conducted a residual analysis:  
- Students who significantly overperformed or underperformed relative to their cognitive ability (based on residuals) were grouped.  
- DESSA scores were significantly higher among overperformers, suggesting that strong social-emotional skills may support achievement even when cognitive ability is lower.

## Implications

This analysis supports the value of including whole-child data—cognitive and social-emotional measures—when identifying students for intervention. Machine learning can enhance this process by uncovering nuanced patterns and providing interpretable insights (e.g., via feature importance and residual analysis) that inform practice.

## Files Submitted

- Jupyter Notebook with full code and results  
- Summary report (this document)