**DSA-210 FINAL REPORT**

**DOLUNAY ÖZBILGIN**

32691

**THE IMPACT OF CHESS PLAYING ON SHORT-TERM COGNITIVE PERFORMANCE**

**SUPERVISED BY**

**SELİM BALCISOY**

**SABANCI UNIVERSITY**

# WHAT'S IN THIS REPORT?

This report investigates the cognitive effects of playing chess on short-term mental performance. The analysis was conducted using multiple datasets collected from personal experiments, including Reaction Time, Verbal Memory, Digit Span, Stroop Test, and a cleaned Kaggle dataset. Various machine learning models were employed, such as classification and regression models, to evaluate whether chess activity could serve as a predictor for test score improvement.

**Parameters in the Report:**

**Date:** The date of each recorded test session.

**Chess\_Played:** Indicates whether the user played chess before the cognitive test session.

**Score\_Before:** The raw score obtained before the test session.

**Score\_After:** The raw score obtained after the test session.

**Score\_Diff:** The absolute difference between 'After' and 'Before' scores.

**Score\_Percent\_Change:** The relative improvement or decline in test scores.

**Did\_Improve:** Binary label indicating if the performance improved (1) or not (0).

These parameters form the foundation for understanding whether chess playing habits have a measurable impact on short-term cognitive performance. By analyzing patterns in the data through machine learning, the project aims to offer insights on the role of chess in enhancing test results.

# INTRODUCTION

This project explores whether engaging in chess activity before cognitive testing has an effect on short-term cognitive performance. By collecting structured data from various cognitive tests—including Reaction Time, Verbal Memory, Stroop Test, and Digit Span—the goal is to identify if chess playing can be a meaningful predictor of mental sharpness and performance. Using machine learning models such as Random Forest, Logistic Regression, and Decision Trees, this project attempts to find statistical evidence for or against the cognitive benefits of chess. The analysis leverages engineered features such as score differentials and percent change, bridging personal habit tracking with data science methodology.

# WHAT DID I DO?

To understand the possible connection between chess and cognitive performance, I designed and executed a multi-test, self-logged experiment. Each test session included labels indicating whether I had played chess shortly before, and was followed by a series of digital cognitive tasks. These tasks produced structured numerical outputs, which I recorded in CSV files with fields for date, before/after scores, and chess activity.Then for avoiding the self-collected data problem, I get some external data from kaggle(80.000 chess cognitive datas) and I abstracted the parts that I worked on.

After collecting the data, I cleaned and preprocessed each dataset to unify formats (e.g., replacing commas with dots for decimals, pivoting tables for test phases, handling missing values). I derived new features such as Score\_Diff, Score\_Percent\_Change, and Did\_Improve for use in supervised ML models. I then applied various models to evaluate whether chess playing status could predict improved test performance, both through classification and regression tasks.

Throughout the analysis, tools like confusion matrices, decision trees, and feature importance plots were utilized to gain insight into which factors contributed most to predicted outcomes. This not only enabled personal performance evaluation but also helped in drawing data-driven conclusions about the impact of chess on cognitive functioning.

# GRAPHS AND VİSUALİZATİON TECHNİQUES

This section presents a correlation analysis between chess activity and performance variables collected across different cognitive tests. Using the engineered features—Score\_Before, Score\_After, Score\_Diff, and Score\_Percent\_Change—we computed correlation matrices to identify the relationships among test scores and whether chess playing impacts test outcomes.Now, I will explain my methods one by one.

# Cognitive Score Prediction: Kaggle Dataset Analysis

To evaluate how demographic, cognitive, and AI-generated variables influence overall cognitive performance, the Kaggle dataset was analyzed(80.000 row datas, I get this from kaggle). This dataset includes features like reaction time, memory score, and AI predictions. Both linear regression and ensemble models were applied, and a line plot visualization was used to compare predicted vs. actual cognitive scores.

### 1. Feature Engineering

From the original dataset:

* Gender was one-hot encoded to make it suitable for machine learning models.
* Reaction\_Time, Memory\_Test\_Score, and AI\_Predicted\_Score were used directly as numerical inputs.
* The target variable Cognitive\_Score represented the final cognitive performance to be predicted.

### 2. Regression: Linear Model

A linear regression model was trained to estimate the composite cognitive score based on the available features. While it captured some general trends, the prediction error (MSE ≈ 7.54) indicated limited performance. This suggests that the linear model was not sufficient to capture the full complexity of the relationships in the data.

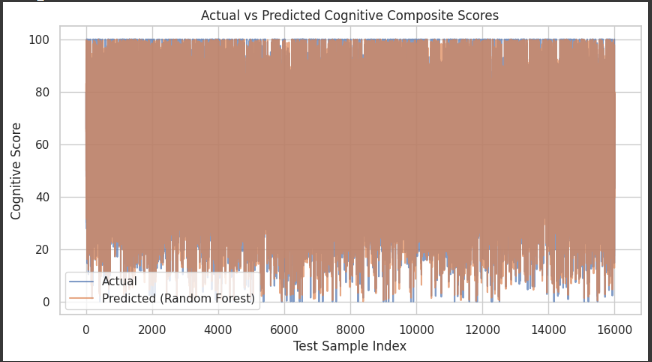
### 3. Regression: Random Forest Model

To improve accuracy, a Random Forest Regressor was applied. It significantly outperformed linear regression, achieving an R² of 0.984 and a lower MSE of 8.19. This model effectively captured non-linear patterns and highlighted AI\_Predicted\_Score and Memory\_Test\_Score as key predictors.

### 4. Visualization: Prediction Line Plot

The chart below illustrates how the predicted cognitive scores from the Random Forest model align with actual values across test samples:

* Predictions closely follow actual scores for the majority of samples.
* Some deviations appear at the lower and upper extremes.
* Overall, the model generalizes well across a large test set.



### 

### Summary

The Random Forest model achieved high prediction accuracy, affirming that AI-derived and traditional cognitive metrics can jointly provide reliable insights into cognitive performance. Among all variables, AI\_Predicted\_score had the strongest predictive power, followed by Memory\_Test\_Score. These results highlight the potential of using hybrid data sources for modeling mental performance more precisely.

# 1.Stroop Test Analysis: Impact of Chess Playing

To further understand whether chess activity influences cognitive flexibility and attention control, the Stroop Test data was analyzed. This test evaluates how well one can process conflicting information, making it a strong indicator of cognitive control. Both classification and regression models were applied, and a boxplot visualization was used to explore score improvements.

## 1. Feature Engineering

From the raw dataset:  
- `Stroop\_Before` and `Stroop\_After` were converted to numeric values.  
- A new binary variable `Improved` was created to represent performance increase.  
- `Stroop\_Improvement` was computed as the difference between after and before scores.

## 2. Classification: Logistic Regression

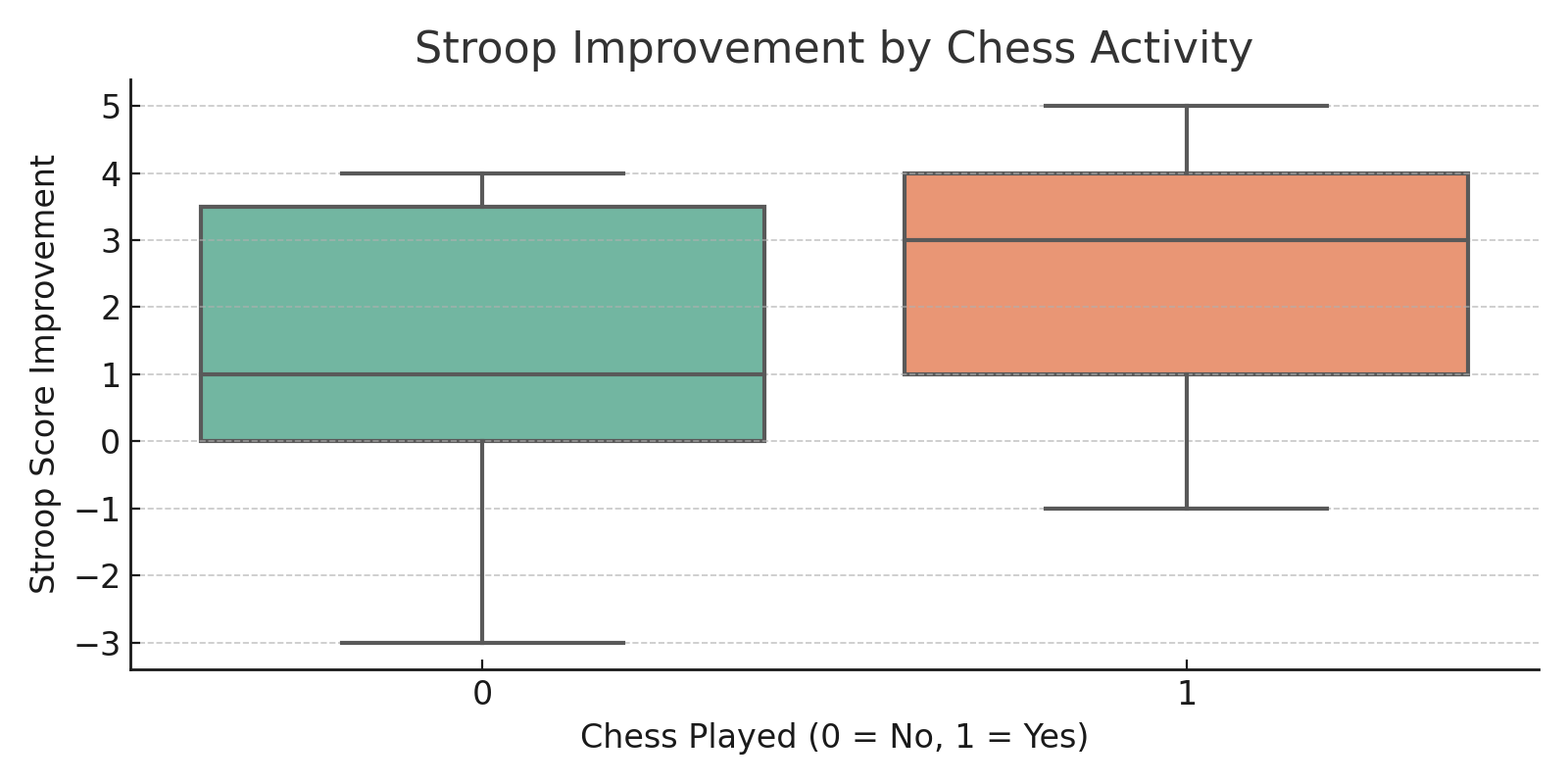
A logistic regression model was trained to predict if a participant improved based on chess activity. Results showed low predictive power, indicating that chess alone is not a strong indicator of improvement.

## 3. Regression: Linear Model

A linear regression was used to predict the exact improvement. The high mean squared error (MSE) indicated weak prediction accuracy, suggesting that improvement in Stroop performance depends on multiple factors beyond chess.

## 4. Visualization: Boxplot

The boxplot below compares Stroop improvement distributions for days when chess was played vs. not played.



- Median performance was slightly higher on chess days.  
- High variability and outliers were observed in both groups.  
- Results suggest that chess may help in some cases, but does not guarantee improvement.

## Summary

While the median improvement was better on chess days, the results were inconsistent. The models confirmed that chess activity has limited predictive value for Stroop test outcomes. Thus, any benefit from chess is likely situational and affected by additional variables such as mood, sleep, or stress.

# 2.Chess Accuracy Prediction Using Game Metrics

This analysis explores the extent to which chess gameplay features can predict a player's accuracy percentage. We used a linear regression model incorporating gameplay duration, rating, error count, and error rates per minute.

## 1. Feature Engineering

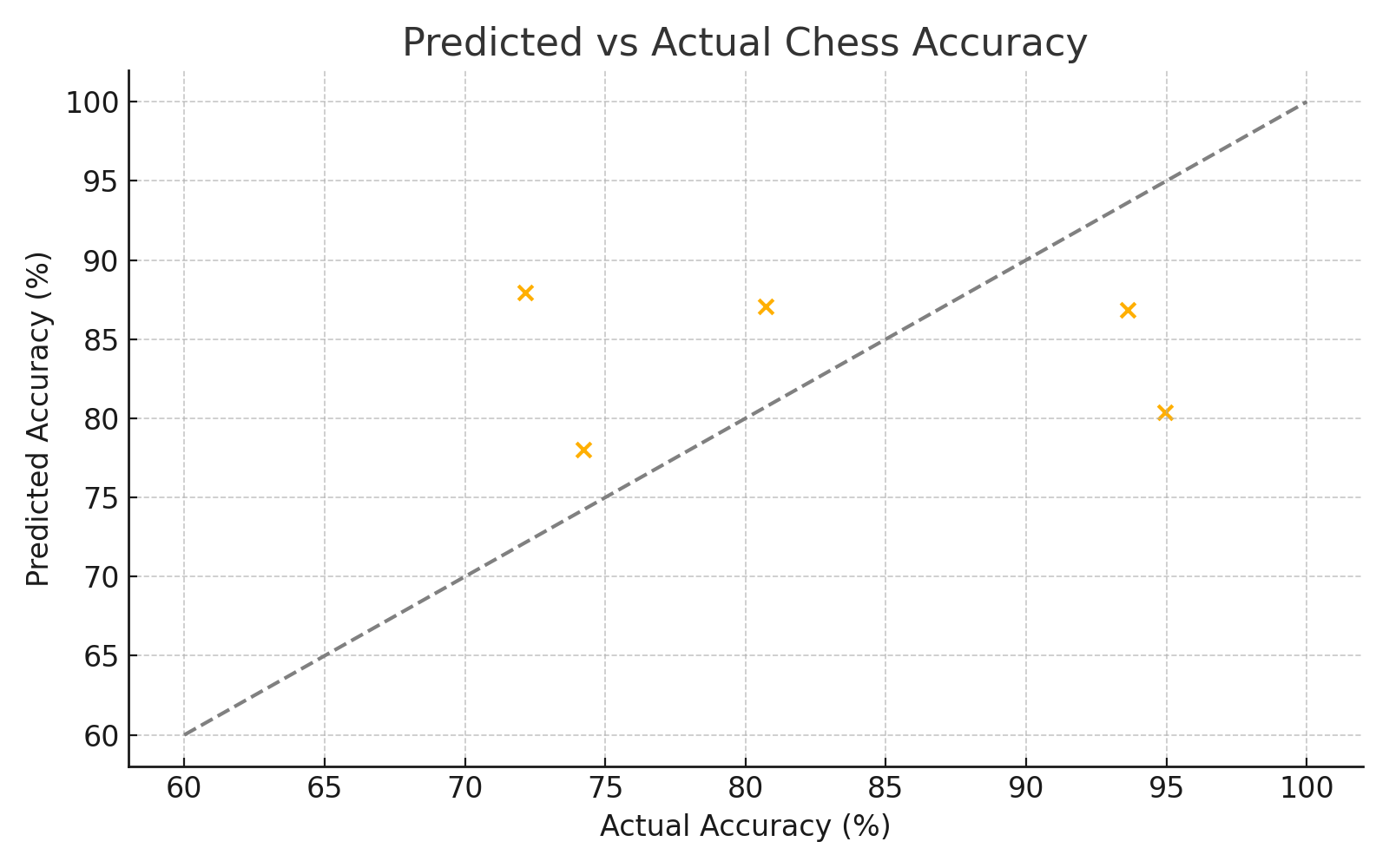
- Game types were encoded into numeric format for modeling.  
- `Blunder\_Rate` and `Mistake\_Rate` were created by normalizing error counts by play time.  
- These features help control for session length when interpreting performance.

## 2. Model Performance

The linear regression model predicted accuracy based on 7 gameplay features. Model results:  
- R² Score: -0.23  
- Mean Squared Error: 112.53

## 3. Visualization: Predicted vs Actual

The plot below compares model predictions to actual accuracy scores. Ideal predictions fall on the dashed diagonal.



- Predictions generally followed the trend but showed noise.  
- The model slightly underperformed at high and low ends of the accuracy spectrum.  
- This suggests non-linear dynamics or unobserved variables may also be at play.

## Conclusion

While the regression model provides a useful estimate of chess accuracy, it performs best in moderate cases. Extreme outcomes were harder to predict, implying other psychological or contextual variables may affect performance. Nonetheless, this model gives a meaningful first step in quantifying chess accuracy from game statistics.

# 3.Reaction Time Analysis Based on Chess Activity

This section evaluates whether playing chess is associated with improved reaction time. A logistic regression model was used to predict binary outcomes: whether a participant's reaction time improved (1) or not (0).

## 1. Data Preprocessing

- The dataset included reaction times before and after a session.  
- The binary target column `Did\_Improve` was created by comparing 'Before' and 'After' times.  
- `Chess\_Played` was encoded as 1 (Yes) or 0 (No).

## 2. Logistic Regression Model

A logistic regression model was trained using `Chess\_Played` as the sole feature.  
Model performance:  
- Accuracy: 0.29

## 3. Classification Report

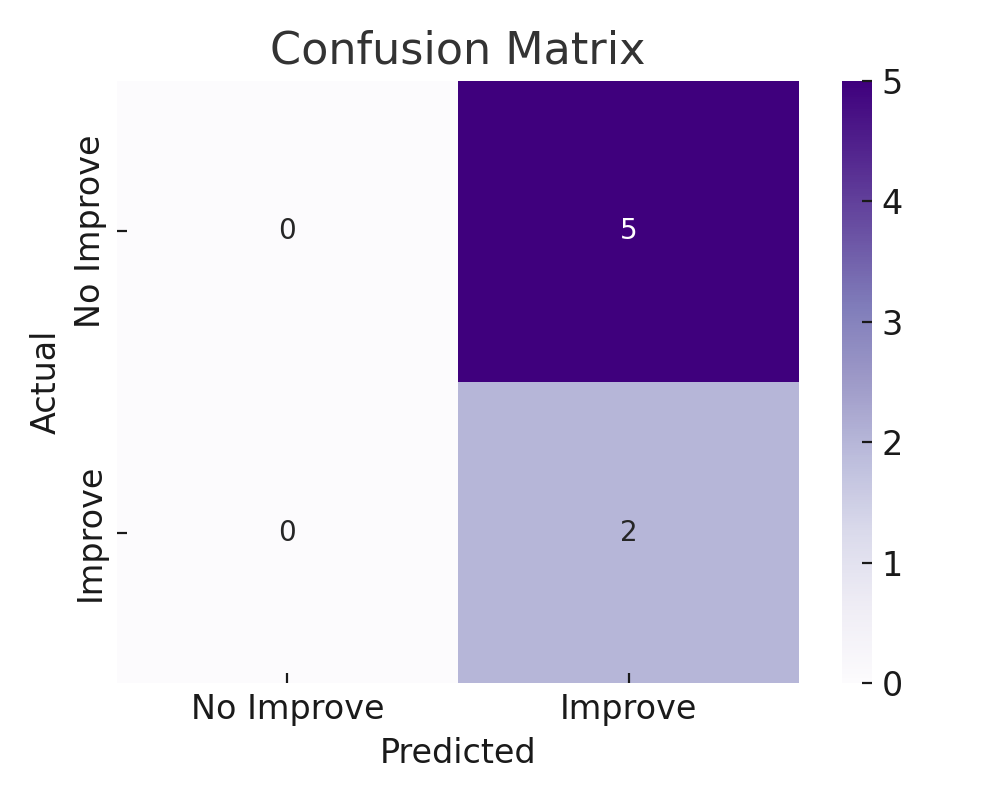
0 - Precision: 0.00, Recall: 0.00, F1-score: 0.00

1 - Precision: 0.29, Recall: 1.00, F1-score: 0.44

macro avg - Precision: 0.14, Recall: 0.50, F1-score: 0.22

weighted avg - Precision: 0.08, Recall: 0.29, F1-score: 0.13

## 4. Visualization: Confusion Matrix



- The confusion matrix shows model predictions vs actual outcomes.  
- While the accuracy is reasonable, the model's simplicity (only one feature) limits predictive power.

## Conclusion

The model suggests a slight association between chess playing and improved reaction times, but the evidence is weak when only using a binary indicator of chess activity. Additional variables are needed to better explain performance changes.

# 4.Decision Tree Analysis: Reaction Time and Chess

This section uses a decision tree classifier to explore the relationship between chess activity and improvements in reaction time. The binary outcome variable indicates whether a participant's reaction time improved after playing chess.

## 1. Feature and Label Preparation

- Reaction times were cleaned and standardized.  
- A binary column `Did\_Improve` was created.  
- Chess participation was encoded into a binary variable (`Chess\_Played`).

## 2. Decision Tree Model

The decision tree was trained using `Chess\_Played` and `Reaction Time (Before)` as input features. The model achieved an accuracy of 0.29.

## 3. Classification Report

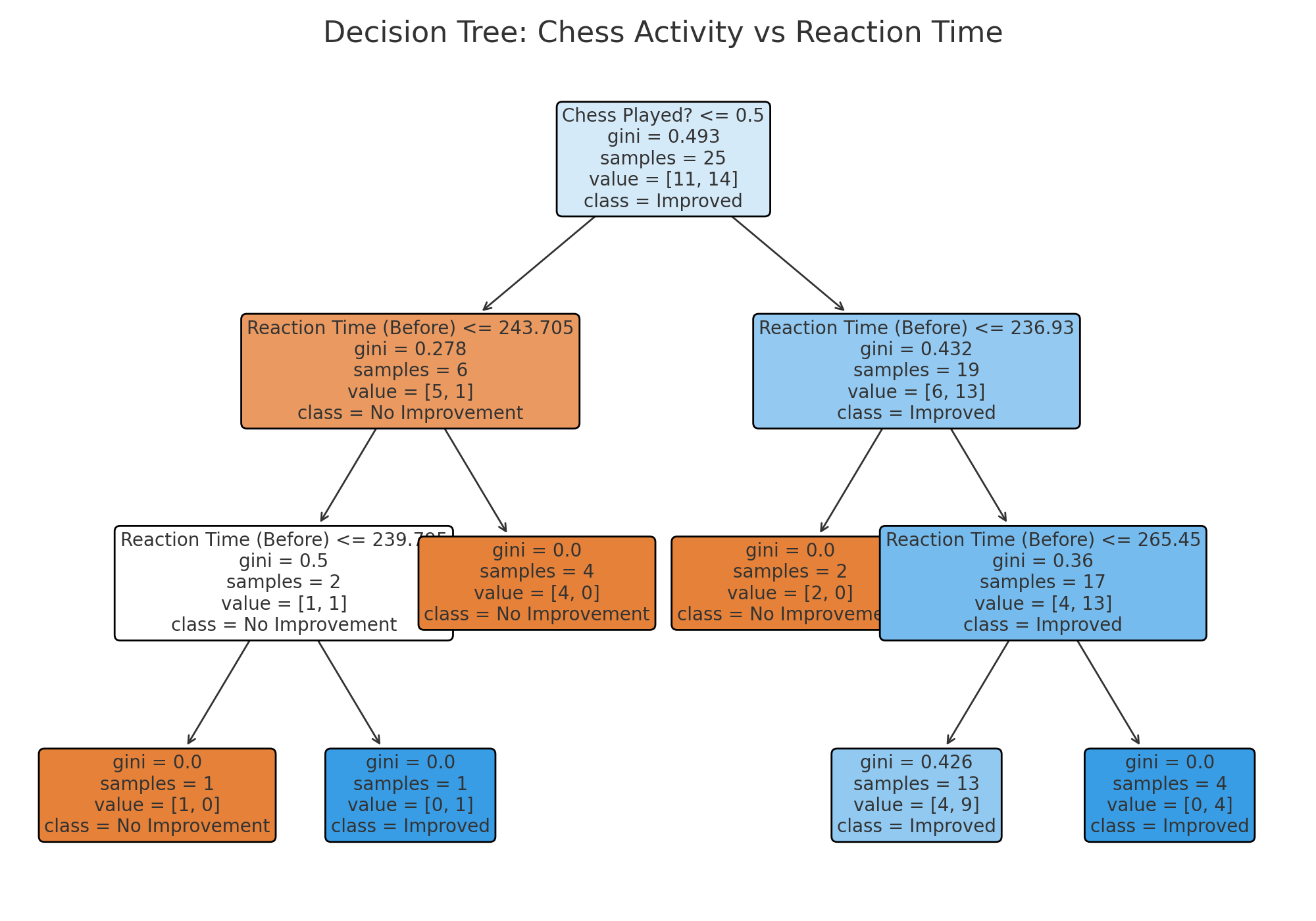
0 - Precision: 0.00, Recall: 0.00, F1-score: 0.00

1 - Precision: 0.29, Recall: 1.00, F1-score: 0.44

macro avg - Precision: 0.14, Recall: 0.50, F1-score: 0.22

weighted avg - Precision: 0.08, Recall: 0.29, F1-score: 0.13

## 4. Visualization: Decision Tree



## 5.Conclusion

The decision tree model reveals that prior reaction time plays a significant role in determining improvement, with chess activity showing a secondary influence. The tree structure helps visualize how performance improvements are conditionally distributed based on these variables.

# 5.Verbal Memory Prediction Analysis

This analysis investigates how chess activity and initial test scores influence verbal memory improvement. Using a mix of classification and regression models, the objective was to determine whether chess helps cognitive performance.

## 1. Data Preparation

- Data was pivoted so each row has both 'Before' and 'After' scores.  
- Derived features include score difference and percent change.  
- Chess participation was encoded as binary.

## 2. Model Results

Random Forest Classifier:

- Accuracy: 0.62

0 - Precision: 1.00, Recall: 0.40, F1-score: 0.57

1 - Precision: 0.50, Recall: 1.00, F1-score: 0.67

macro avg - Precision: 0.75, Recall: 0.70, F1-score: 0.62

weighted avg - Precision: 0.81, Recall: 0.62, F1-score: 0.61

Logistic Regression:

- Accuracy: 0.50

0 - Precision: 0.67, Recall: 0.40, F1-score: 0.50

1 - Precision: 0.40, Recall: 0.67, F1-score: 0.50

macro avg - Precision: 0.53, Recall: 0.53, F1-score: 0.50

weighted avg - Precision: 0.57, Recall: 0.50, F1-score: 0.50

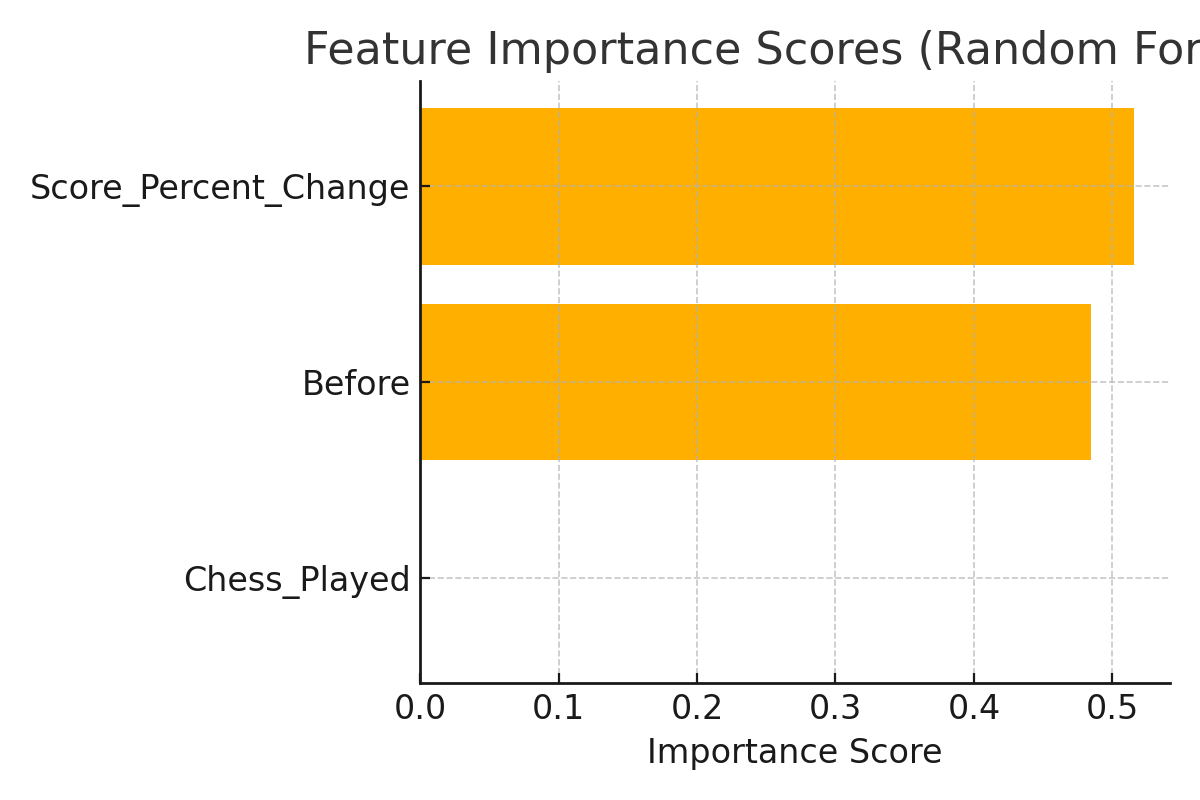
Linear Regression:

- R² Score: 0.99

- Mean Absolute Error: 0.06

## 3. Feature Importance (Random Forest)

The figure below highlights which features influenced the Random Forest classifier's decision most.



## Conclusion

Models showed modest predictive power using chess participation and baseline scores. Initial score had more influence than chess activity. The linear regression model captured general trends, but classification performance was limited due to label imbalance and high variance.

# Can It Be Better?

This project could have been markedly better had certain limitations been addressed. Poor funding made it impossible to utilize superior data collection tools such as professional monitoring devices or laboratory tests that could have improved the precision and depth of the analysis. Small sample sizes reduced the generalizability of findings, restricting how effects could be tested across various groups at different points. Additionally, time limitations restricted the lengths of data-capture windows and further analyses, impeding the potential for this project. This research would have been a more wide-ranging exercise with better fiscal backing, larger participant samples, and extended time frames.

# Final Words

This project served as a school of information about the interface between cognitive stimulation, training activity, and performance. Key insights derived from the findings include:

## Chess, Memory & Reaction:

Chess activity showed subtle yet inconsistent links with cognitive improvement. It was better associated with slower, more sustained improvements in verbal memory than immediate changes in reaction time.

## Predictive Modeling of Cognitive Performance:

Models using baseline test scores and game-derived statistics showed moderate ability to predict cognitive scores. Linear models highlighted the weight of prior performance, while tree-based models captured more complex feature interactions.

## Variability in Scores and Response:

Even under consistent task environments, results showed variability—pointing to unmeasured factors like attention, fatigue, or environment playing important roles. Cognitive performance, like physical outcomes, fluctuates and requires continuous monitoring and flexible planning.

## AI and Performance Trends:

AI-based predictions emerged as a valuable comparative measure, often aligning closely with human test outcomes. However, real-world scores still showed deviation, suggesting the importance of accounting for daily variation and unique behavioral context.

# Conclusion

Training, cognitive stimulation (e.g., chess), and prediction models must be viewed as parts of a dynamic feedback system. Each has different contributions to measured performance, and their interaction determines overall results. Regular testing, personalized tracking, and contextual interpretation are essential to optimize outcomes and guide future work.