1. Introduction

This document evaluates the accuracy of the machine learning models used to predict winning decks in Magic: The Gathering (MTG) Commander tournaments. Two models were implemented: a Regression Random Forest to predict a deck's relative placement in the tournament (RankPercentage) and a Classification Random Forest to predict a deck's RankBracket. This assessment explores their effectiveness, limitations, and areas for improvement.

2. Accuracy Metrics and Evaluation

Regression Model

- **Root Mean Squared Error (RMSE):** Measures the average prediction error. A lower RMSE indicates better accuracy.
- **Mean Squared Error (MSE):** Averages the squared differences between predicted and actual RankPercentage values.
- **Mean Absolute Deviation (MAD):** Evaluates how much predicted values deviate from actual results.
- **R^2 Score:** Represents how well the model explains variance in the data (closer to 100% is ideal).

Latest Results:

- RMSE: 0.2786
- MSE: 0.0776
- MAD: 0.2342
- R^2 Score: 6.14% (suggests weak predictive power)

Classification Model

- **Accuracy:** Measures the percentage of correctly predicted RankBrackets.
- **Error Count:** The number of incorrect predictions.

Latest Results:

- Accuracy: 39.61%
- Error Count: 1758

3. Observations and Limitations

Through testing on 11 different tournaments, several patterns emerged:

- The model consistently placed the same commanders near the top across multiple tournaments.

- Despite a deck's average RankPercentage prediction, small scoring variations caused some decks to drop multiple places in the final ranking.
- Most decks maintained similar scores across different tournaments, suggesting that the model captures core deck strengths but struggles with fine distinctions.
- Performance Differences: The model appears to perform better in smaller tournaments than in larger ones, possibly due to reduced variance and competition dynamics.

Despite strong feature engineering, both models face challenges:

- **Feature Overlap:** Many decks share similar features, making differentiation difficult.
- **Tournament Structure:** The model does not yet consider matchups, player behavior, or turn-based results.
- **Feature Correlation Issues:** Some features, such as Commander Diversity and Mana Curve Smoothness, show weak correlation with RankPercentage.
- **Class Imbalance:** The classification model predicts most decks as RankBracket 4 due to limited variation in training data.

4. Real-World Significance of Accuracy

At current accuracy levels, the model is useful for general insights rather than precise tournament predictions. While not perfect, it provides a structured approach to evaluating deck strength. Higher accuracy would be needed for professional tournament applications.

5. Potential Next Steps

To improve accuracy, future iterations should include:

- **Expanded Feature Set:** Include match-up data, win/loss tracking, and player behavior.
- **Alternative Algorithms:** Explore ensembling (e.g., stacking models) or neural networks.
- **Data Augmentation:** Increase training data with simulated tournaments.
- **Further Hyperparameter Optimization:** Extend Bayesian Optimization for better tuning.

6. Conclusion

While the models provide insightful predictions, there is room for improvement. By refining features and optimizing algorithms, the accuracy of tournament outcome predictions can be significantly enhanced in future iterations.