Hypothesis Analysis: Evaluating Predictive Features in MTG Tournament Outcomes

Introduction

The primary hypothesis of this project was that certain measurable deck features—such as keyword synergy, mana curve smoothness, tutor count, and commander diversity—would correlate strongly with a deck's success in MTG Commander tournaments. However, through extensive analysis and visualization, it became evident that while some features influence rankings, others lacked strong predictive power. This section presents an evaluation of the hypothesis using heatmaps and scatterplots to assess feature importance and suggest potential improvements.

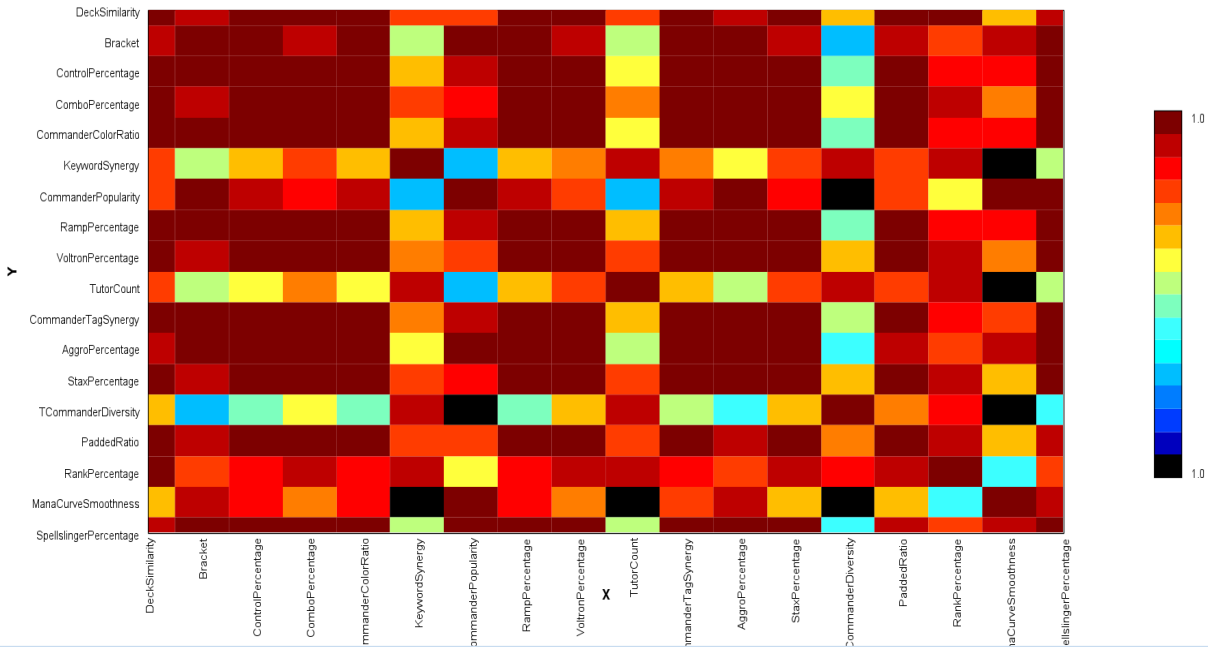
---

1. Regression Heat Map: Feature Correlation Analysis

The regression heatmap revealed that most features had strong correlations with one another, except for a few outliers:

- Weakly correlated features:
  - KeywordSynergy
  - TutorCount
  - CommanderDiversity (least correlated)
  - ManaCurveSmoothness

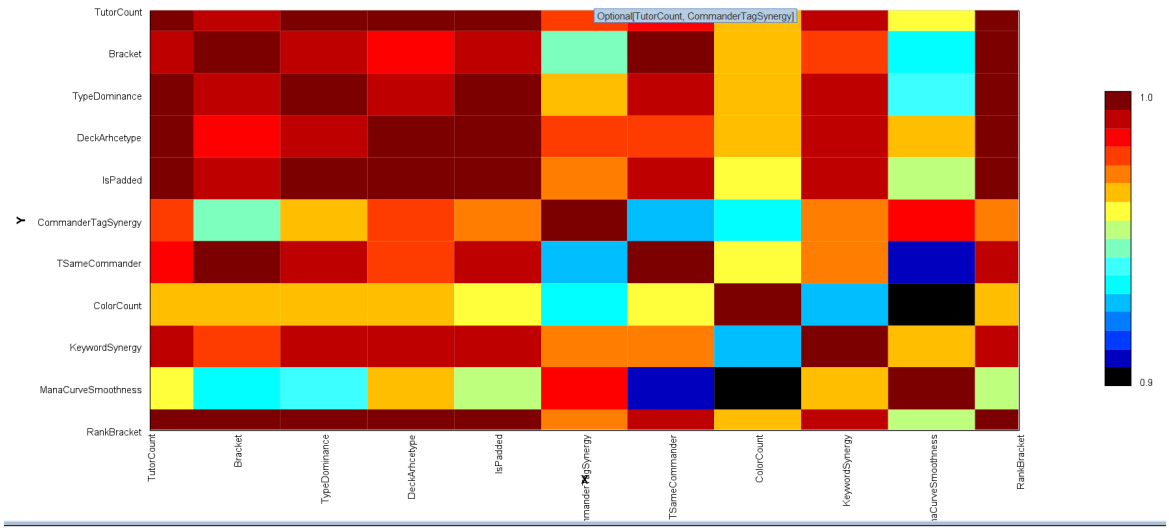
The strong internal correlations between most features indicate redundancy, meaning multiple features may be capturing similar information. However, Commander Diversity's weak correlation suggests that, within this dataset, a tournament's variety of commanders does not significantly influence a deck's likelihood of winning.



2. Classification Heat Map: Low Predictive Strength

Unlike the regression heatmap, the classification heatmap displayed mostly weak correlations, with the majority of the graph being yellow and blue, signifying low statistical relationships. Only a small subset of features had moderate correlations.

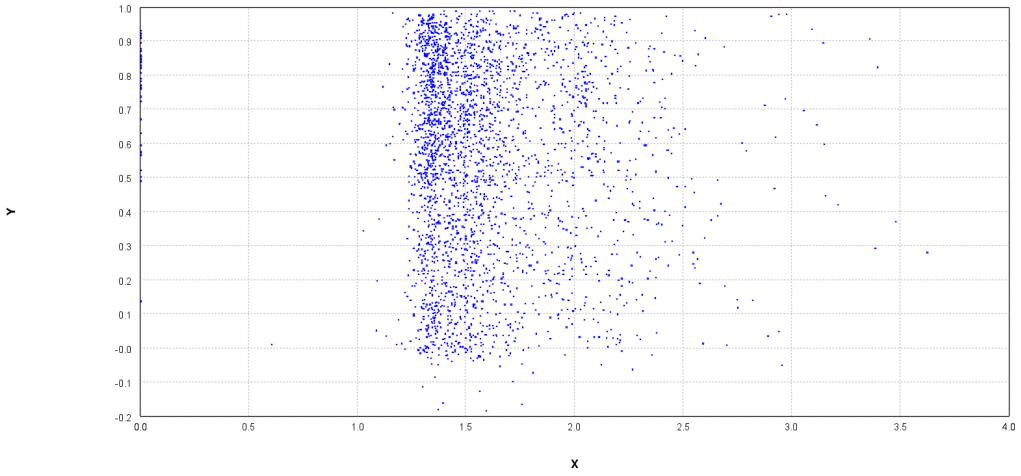
- This suggests that the current classification features are not sufficiently distinct to classify decks effectively into ranking brackets.
- Additional contextual features—such as player behavior, game history, and deck matchups—may be necessary to improve classification accuracy.



3. Scatterplot Analyses: Feature Distribution Patterns

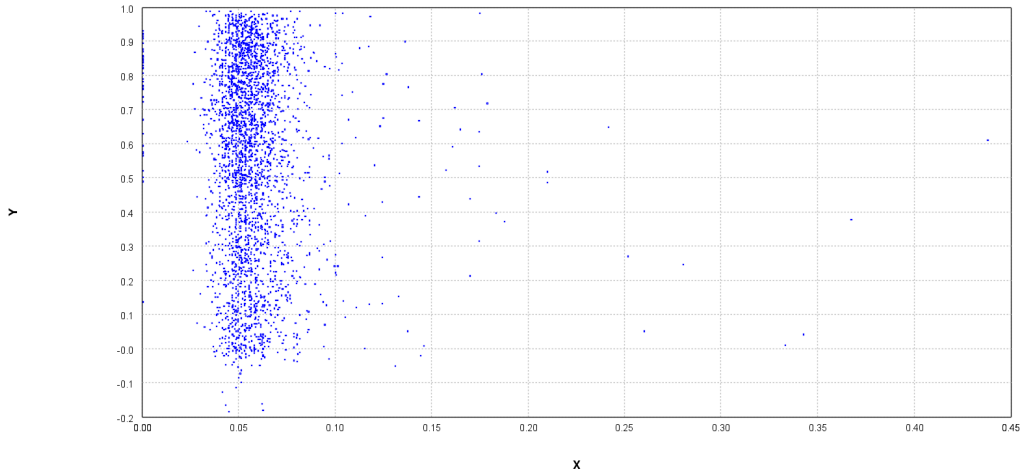
(a) ManaCurveSmoothness vs. RankPercentage

- The scatterplot showed that most decks had values between 1.0 and 1.5, indicating minimal variation in this feature.
- A high degree of clustering suggests that decks are built too similarly, making it difficult for this feature alone to determine winning decks.



(b) KeywordSynergy vs. RankPercentage

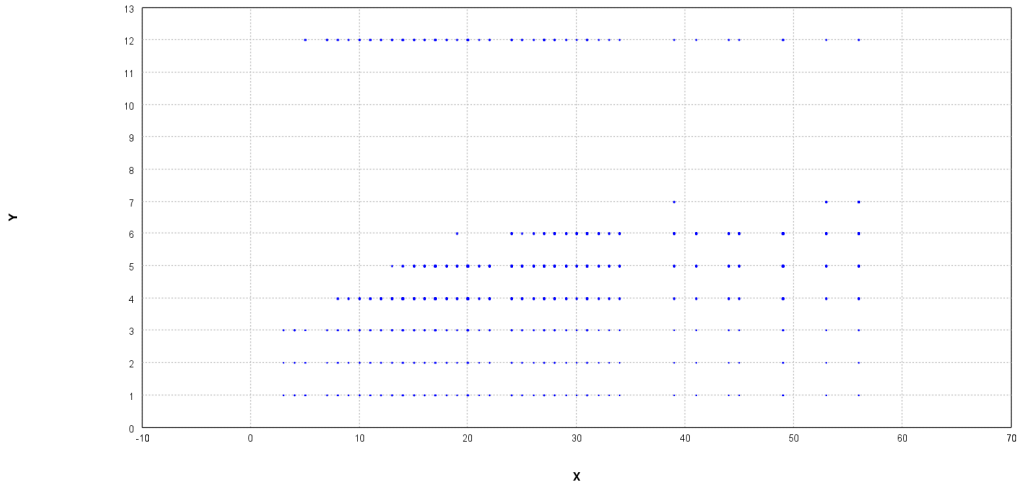
- Almost all decks had a KeywordSynergy value between 0.0 and 0.1.
- The lack of distribution in this feature makes it a poor predictor of tournament success.
- A possible improvement could be utilizing EDHREC synergy scores, which account for card combinations across millions of decks.



(c) TSameCommander vs. RankBracket

- This scatterplot revealed a notable inverse relationship: the more common a commander, the lower its ranking.

- While common commanders may indicate strong meta choices, overrepresentation may lead to “counters” being more widely prepared, reducing their overall success.
- This suggests that less predictable commander choices might have a competitive edge in certain tournament environments.



Conclusion and Future Improvements

The hypothesis that certain deck features strongly correlate with tournament success is partially supported. While some features show moderate predictive power, others exhibit weak correlations and require refinement.

Key Takeaways:

- Feature redundancy exists—many features correlate strongly, meaning some could be removed without losing information.
- Commander Diversity and Keyword Synergy were weaker predictors than expected.
- Common commanders tend to underperform in tournaments, indicating a potential advantage in less common commander choices.
- Future improvements should include:
    - Tracking match history to incorporate opponent interactions and win/loss patterns.
    - Integrating turn data to understand how quickly decks achieve victory.
    - Enhancing synergy calculations using external databases like EDHREC.

These findings provide valuable insights into the nuances of deck performance in competitive MTG tournaments and suggest several avenues for enhancing the predictive accuracy of the model.