

# Developing a Regression Algorithm for Predicting Magic: The Gathering Card Efficiency in Draft Format

Jonathan Tuck

**Abstract**—This paper seeks to predict efficiency ratings for Magic: The Gathering cards in the Draft format. This paper first uses reliable data gathered from the Magic community as the true efficiency ratings. Then, this paper discusses which of the available features are used and the reasoning behind using or not using a feature. This paper then examines a variety of regression algorithms to determine the highest accuracy algorithm when tested on the test data. We obtain results that conclude that Multivariate Adaptive Regression Splines with thresholding yields the most accurate results on the test data. Finally, this paper discusses potential improvements and extensions to this work.

## I. INTRODUCTION

Magic: the Gathering (Magic) is a trading card game that, since 1993, has controlled the genre with over 20,000,000 players. There are many formats to uniquely play Magic; one of the most played formats being Draft, where players open a pack, pick a card out of the pack, pass the remaining cards to a person next to them, and repeat this process until three packs (typically all of the same expansion) have been opened and distributed this way. With over 70 unique expansions, the Draft format varies in both complexity and entertainment value every few months.

In Draft, one of the most important factors in building a winning deck is drafting "efficient" cards, or cards that are relatively cheap (in terms of its mana cost) for the cumulative effect of the card on the game. For the purposes of being complete, we define card efficiency as follows:

$$\text{Card Efficiency} = \frac{\text{Overall Positive Effect of Card}}{\text{Mana Cost}}.$$

This project seeks to predict a card's efficiency in a one-expansion draft format. This problem is non-trivial, as the power of a card is not readily or easily noticable; it is only when compared to the rest of the expansion for both power anomalies and synergies with other cards in the expansion that card efficiency can be realized. Ultimately, the motivation behind this project is to contribute to the development of a system to train Magic players in drafting efficient cards.

## II. DATASET AND RELATED WORK

With over 70 unique expansions containing between 90 and 450 cards each, there is an abundance of data on Magic cards. In addition, the Magic community is a vibrant one, including one well-used technical analysis of Magic cards.

This project's data was gathered from DraftSim, the leading Magic Draft format simulator [3]. DraftSim is used by the entire Magic community, including its professionals, and is typically considered "correct" (a technical term used by the Magic community) when making decisions on Draft. For this reason, we use DraftSim's data as the "true" efficiency rating in classifying cards.

DraftSim ratings range from 0.0 to 5.0, in 0.1 increments. Data was gathered from 20 of the most recent expansions and will be used in this experiment. Ultimately, however, only the previous expansion will be used to predict the ratings of the test data, the current expansion (discussed further in section V), as it is closely related in both power level and similarity of mechanics. The data gathered from DraftSim includes a name and a corresponding rating as described above [3]. Based on the data from [3], this project is modeled as a regression problem.

## III. ANATOMY OF A MAGIC CARD

The anatomy of a Magic card is fairly straightforward for all cards. Each card has a multitude of features that in combination uniquely characterize itself. These features include the card name, the mana cost of the card, the artwork, the expansion symbol, the unique on-card effect, the power/toughness (for creature cards only), the card number, rarity, & illustrator, the flavor text (text not affecting the game state), and the card type and subtype.

Within the card type there exists two supertypes: the "creature" spell, and the noncreature spell. Within the noncreature spell supertype includes the "sorcery" (a spell you can only play on your own turn), the "instant" (a spell you can play on anybody's turn), the "artifact" (a spell, sometimes also a creature, that stays on the game board until destroyed), the "enchantment" (similar to an artifact but is "attached" to a particular creature or player), and the "plainswalker" (a "hero" card that can drastically change the game state in favor of its controller.) In addition, there is a third supertype, the "land," which is the main resource for paying mana costs of cards. Lands are typically given to players after drafting and players can select as many of lands in any combination as needed, so the availability of lands very rarely, if at all, has any influence on drafting. This project therefore does not consider lands in its analysis.

## IV. FEATURES

As discussed in the above section, a card includes a various number of features, many (but not all) of which are important



Fig. 1. A complex Magic card, "Emrakul, the Promised End," with explanation of its Magic card features.

in determining both the overall positive effect of a card as well as its mana cost.

Figure 1 is an example of one of the relatively more complex cards created [1], with annotations illustrating the various potential features of the card. As is hopefully illustrated, it is non-trivial to quantify the efficiency of a given card (for example, one might ask if the cumulative effect of this card on any single match efficient for a mana cost of 13 colorless mana, or if there exists another card in this expansion that provides similar or identical effects for a lesser mana cost.)

In deciding the correct classification algorithm for this data, it was necessary to *a priori* decide if some of a card's features would not have an effect on a card's efficiency. In particular, it was decided that the card name, the flavor text, the artwork, the card number, expansion, rarity, and illustrator were unnecessary in deciding a card's efficiency, as the above features are purely aesthetic.

The important features of a card that this project focuses on include the mana cost, the unique on-card effects, the card type and subtype, and the power and toughness (if it exists, as it does not exist for noncreature spells).

### A. Mana Cost

At first glance, it looks as though mana cost is straightforward to model, but it is important to realize that the importance of mana cost is not simply in the number. There are six types of mana: white, blue, black, red, green, and colorless (for reference, Figure 1 only requires colorless mana to be used.) Any color of mana can be used to pay for colorless mana. This is important as the game rules dictate that you can only add one type of mana to your mana "pool" per turn. This means that if one has a choice of two cards that perform the same

function but the first card costs four colorless mana and second card costs four red mana, then the first card is objectively more efficient than the second card. Figure 2 is an example of objectivity of efficiency, illustrating the two cards, one objectively better than the other, because of the reduced mana cost.



Fig. 2. A card (right) being compared to a objectively more efficient card than it (left). The two cards perform identical actions, but the leftmost card costs one less colorless mana than the rightmost card.

In addition, some unique on-card effects require additional mana costs, so in reality, there are 12 important mana-dependent features: the six types of mana for both the primary mana cost and the secondary mana cost. There are extremely few cards with tertiary or further mana costs, and none in either the training data or the test data.

### B. Unique On-Card Effects

Typically, on-card effects increase the complexity of the game-state, and are almost always increase the cumulative positive effect of a card. The method by which this project characterizes unique on-card effect features are explained in greater detail in Section V, Subsection A.

### C. Card Type and Subtype

The card type is important for card efficiency, as card types dictate when you can play a card. To illustrate why card type efficiency is important, imagine that there exists two cards, both completely identical except for that one of the cards is a sorcery, and the other an instant. In this scenario, the instant would be considered strictly more efficient, as you can play it at any time in the game, whereas with the sorcery, you would only be able to play it during your turn.

In addition, the card subtype is important for quantifying card efficiency, as many cards are more powerful when combined with cards of the same subtype (in Magic, this is called a "tribal" effect). For example, a subtype in the Eldritch Moon expansion is "Vampire." There exists some cards that increase the power of other Vampires, so we can say that the Vampire subtype is a feature of a card that increases the overall positive effect of that card by some amount (determined by the

classifier, in this project). This logic is applied to all subtypes in an expansion.

#### D. Power and Toughness

The power and toughness are the most straightforward features contributing to the efficiency of a creature card, as they follow the trend that more power and toughness is strictly more efficient. For noncreature spells, this feature is ignored, as these cards do not have power and toughness.

### V. METHODS

#### A. Assumptions

In order to make this problem possible to solve, it is imperative that a few assumptions are made about the data and features.

The most drastic assumption that will be made is in quantifying the unique on-card effect. As demonstrated in Figure 1, quantifying unique on-card effects is only non-trivial but also computationally hard, as many cards have effects that are not seen on any other card (such as the effects seen in Figure 1).

For this reason, a typical Magic paradigm will be used as an assumption in order to model this feature: that the amount of text of the unique on-card effects proportionally corresponds to that card’s power. In explicit terms, the power of the unique on-card effects of a card will be modeled by their length. Of course, there are outliers in this assumption, but they are few. It is not an incorrect assumption to say that the power of the unique on-card effects is typically dependent on the length of the text on that card, as typically a card’s power due to its on-card effect is related to the complexity of the effect on the board state, which typically requires more words to explain.

#### B. Preliminary Data

In order to develop a predictor for this problem, this project shall use one expansion, the “Eldritch Moon” expansion, as the training set. The classifier will be built on Eldritch Moon and then tested on the other expansions. The metric used for classification error will be absolute difference between the Eldritch Moon classifier’s prediction and the true value of the classification gathered from [3].

This project mainly uses two open-source software packages for machine learning: TensorFlow and scikit-learn [4] [5]. These two software packages allow for easy utilization of various machine learning algorithms and analysis.

This project has investigated a multitude of algorithms in order to decide on the best predictor for this particular problem. In particular, the algorithms investigated thus far include:

- Ordinary Least Squares [6]
- Locally Weighted Linear Regression [6]
- Logistic Regression [6]
- Multivariate Adaptive Regression Splines (MARS) [7]
- Locally Estimated Scatterplot Smoothing (LOESS)[8]

The algorithm that was chosen to use as the regression algorithm for this project was the one that yielded the largest percentage of ratings that were closest in value to the true

TABLE I. PERCENTAGE OF PREDICTED CARD RATINGS UNDER A GIVEN ERROR BOUND, FOR THE TESTED REGRESSION ALGORITHMS.

Error	Percentage of Cards with Errors Under: [%]			
	1.0	2.0	3.0	4.0
Ordinary Least Squares	10	23	52	63
Locally Weighted Linear Regression	34	56	74	82
Logistic Regression	50	65	86	92
Multivariate Adaptive Regression Splines (MARS)	88	95	100	100
Locally Estimated Scatterplot Smoothing (LOESS)	16	23	63	73

ratings. This metric is the main metric used to determine a hierarchy of algorithms for this project. Table I is a table highlighting some of these relevant percentages.

As can be seen from Table I, out of the above algorithms, MARS works the best over the “Shadows Over Innistrad” expansion. Therefore, the results shown in section VI are those when MARS has been used. It should also be noted that the “pyearth” Python package was used in conjunction with scikit-learn in order to implement MARS [10].

The results of Table I make sense in terms of their accuracy. Many cards in Magic expansions are made for a wide variety of formats — not just Draft format — so we would expect the efficiency to be nonlinear with respect to the features. This theory would explain why Ordinary Least Squares performed the worst. However, it is expected that many of the Magic cards can be grouped into subgroups whose efficiency can then be linearly or piecewise-linearly approximated, as cards that were developed with the same tier of power and mana cost should intuitively scale with increasing power of the feature set. This theory explains why the best results come from MARS, which essentially can be thought of nonlinear approximating with sums and products of one or more hinge functions (discussed in further detail in the next section.)

#### C. Multivariate Adaptive Regression Splines (MARS)

Recall that with MARS, the predictions built are of the form

$$\hat{y}(x^{(i)}) = \sum_{i=1}^m a_i B_i(x^{(i)})$$

where  $\hat{y} \in \mathbb{R}$ ,  $x^{(i)} \in \mathbb{R}^n$ ,  $\{a_i \in \mathbb{R}, i = 1, \dots, m\}$  are constants and  $\{B_i(x^{(i)}), i = 1, \dots, m\}$  are products of basis functions  $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$  that are of the form:

$$h_i(x) = I\{x \in R_i\}$$

Where  $I$  denotes the indicator function and  $\{R_i\}_{i=1}^m$  are disjoint subregions representing a partition of a domain of interest  $D$  [7]. For this project, the subregions are constructed such that the basis functions are hinge functions, *i.e.*,  $h_i(x) = \max\{0, x - d\}$ , where  $d \in \mathbb{R}^n$  is some fixed vector.

Our true values are only in 0.1 increments, however; therefore, after predicting the value of a card rating, we round the value to the nearest tenth.

#### D. Cross-Validation

In this project, the method of cross-validation used was hold-out cross-validation, with 33% of the data held out. Other methods of cross-validation were considered, such as  $k$ -fold cross-validation and leave-one-out cross-validation, but ultimately hold-out cross-validation was used due to the abundance of data [9].

### VI. RESULTS AND DISCUSSION

Figure 3 shows the results of the regression algorithm used in this project, when tested against the "Shadows Over Innistrad" expansion, using the MARS algorithm with thresholding. The figure is the proportion of cards below a certain absolute error vs. the error in question. The error is taken to be the absolute difference between the prediction and the true value of the card rating. The results qualitatively look like they are doing well in predicting the correct rating.

Quantitatively, the predictor seems to be working well. The mean-square error of the prediction's ratings against that of the true ratings is 0.669 units<sup>2</sup>. For comparison, The mean-square error of the true ratings against that of a zero-vector of the same dimension is 7.649 units<sup>2</sup>. In terms of proportions of accurate predictions, approximately 60% of all cards are predicted correctly with no error. In addition, over 85% of the predicted ratings are within 1.0 rating units of their true value. Lastly, MARS was the only regression algorithm to predict 100% of the cards within 3.0 rating units, and 4.0 rating units, of the true value. This is to be expected, as by construction, the true values of the ratings do have some degree of subjectivity (see Section II).

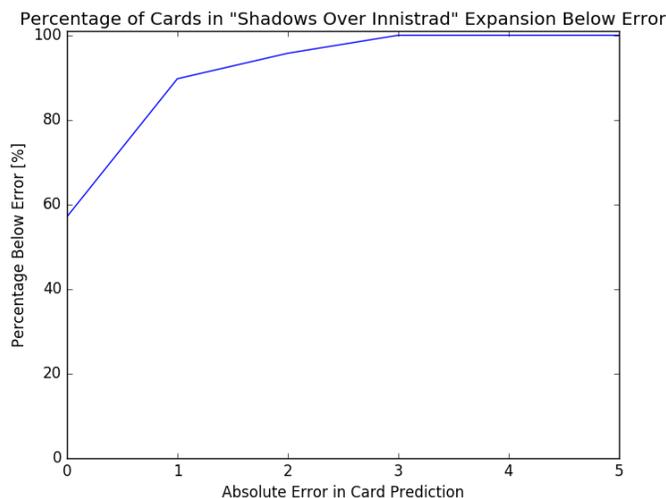


Fig. 3. Proportion of cards below a certain absolute error vs. the error in question, for the "Shadows Over Innistrad" expansion.

### VII. FUTURE WORK

Future work building off of this project is abundant. Naturally, it follows from this project to not only be able to predict the efficiency of a card in a vacuum, but to predict its efficiency

given the past cards a player has drafted in a given draft. This extension would lead to building an efficient, synergistic deck, rather than just determining the efficiency of a single card.

Another natural progression in expanding the scope of this project is to develop an efficiency predictor for cards in other formats of Magic, such as Sealed (constructing a deck from six packs opened at the event), Standard (constructing a deck with the most recent Magic cards from approximately the last two years), Modern (constructing a deck with Magic cards from approximately the last 15 years), Legacy (constructing a deck using any Magic card ever made, minus a list of "banned" cards), and Vintage (constructing a deck using any Magic card ever made, without "bans" on cards).

As the space of available cards (and thus the space of interactions between cards) increases from format to format, the complexity follows exponentially so. In addition, each format has a different base-level of power to be used, so the most powerful card in Draft may not even be good enough to play with in Vintage. The progression from Draft to other formats is not trivial, and will require further investigation.

### VIII. CONCLUSION

This project outlines the methodology and development of a regression algorithm to predict the efficiency of Magic: The Gathering cards in Draft Format. Based on the methodology outlined in this paper, it was empirically found that Multivariate Adaptive Regression Splines (MARS) with hold-out cross-validation on the "Eldritch Moon" expansion training set yielded the smallest overall prediction error when predicting the "Shadows Over Innistrad" expansion ratings.

### REFERENCES

- [1] Wizards of the Coast. *Emrakul, the Promised End*. 2016. [Online]. Available: <http://gatherer.wizards.com/Pages/Card/Details.aspx?multiverseid=414295>. [Accessed: 10 October 2016].
- [2] MTG JSON. 2013. [Online]. Available: <https://mtgjson.com/>. [Accessed: 10 October 2016].
- [3] DraftSim. 2016. [Online]. Available: <http://draftsim.com/>. [Accessed: 1 November 2016].
- [4] TensorFlow. 2016. [Online]. Available: <https://www.tensorflow.org/>. [Accessed: 3 November 2016].
- [5] scikit-learn. 2016. [Online]. Available: <http://scikit-learn.org/stable/>. [Accessed: 9 November 2016].
- [6] Ng, Andrew. *CS 229: Supervised Learning*. 2016. [Online]. Available: <http://cs229.stanford.edu/notes/cs229-notes1.pdf>. [Accessed: 10 November 2016].
- [7] Friedman, Jerome H. *Multivariate Adaptive Regression Splines*. *Ann. Statist.* 19 (1991), no. 1, 1–67. doi:10.1214/aos/1176347963. <http://projecteuclid.org/euclid.aos/1176347963>.
- [8] Cleveland, William S. (1979). *Robust Locally Weighted Regression and Smoothing Scatterplots*. *Journal of the American Statistical Association*. 74 (368): 829–836. doi:10.2307/2286407
- [9] Ng, Andrew. *CS 229: Regularization and Model Selection*. 2016. [Online]. Available: <http://cs229.stanford.edu/notes/cs229-notes5.pdf>. [Accessed: 10 November 2016].
- [10] Rudy, Jason. *A Python implementation of Jerome Friedman's Multivariate Adaptive Regression Splines*. 2016. [Online]. Available: <https://github.com/scikit-learn-contrib/py-earth>. [Accessed: 15 November 2016].