

Color me Mythic

Classifying Magic: The Gathering Cards



Brian Yee

What color is this card?

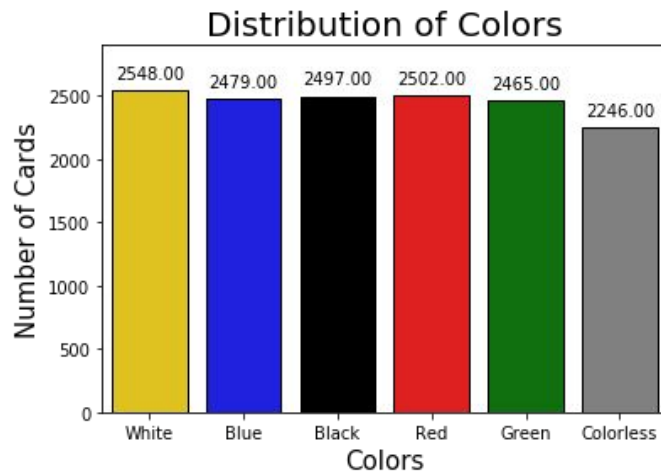
- In Magic: The Gathering, each card can be one of five colors or any combination of these five colors.
 - White, Blue, Black, Red, Green, Colorless
- Each color has its a theme that remains fairly consistent throughout all of Magic history.



The Data

- I got data on every card ever printed by Wizards of the Coast from MTGJSON.com.
- Scraped the MTG wiki for list of standard sets and list of evergreen keywords.

Only included cards that were monocolored and have been in the standard format at some point.

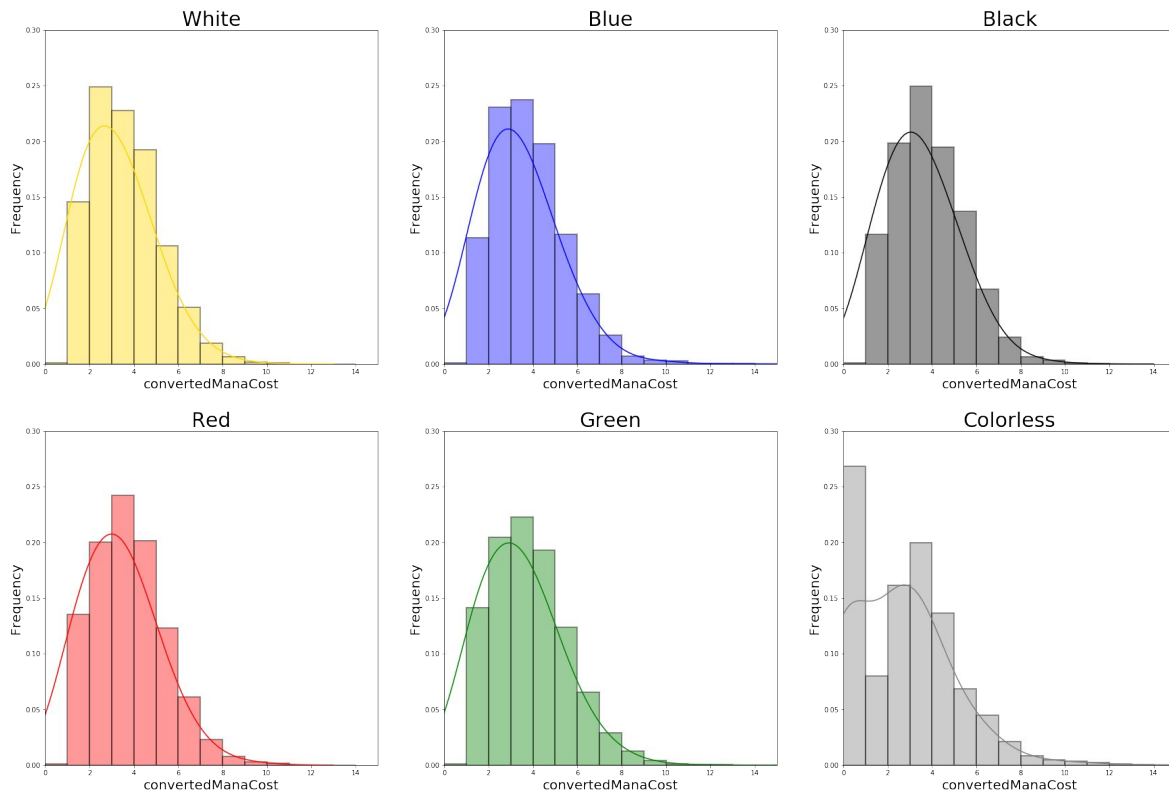


What makes the colors unique?

The converted mana cost of each color varies slightly.

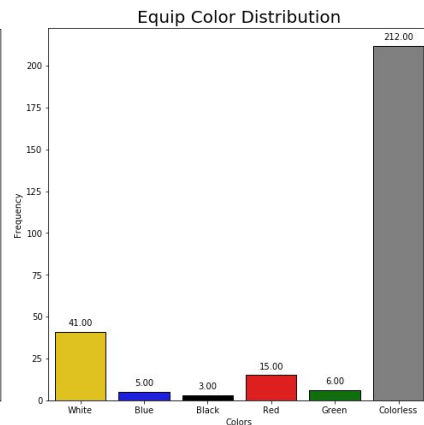
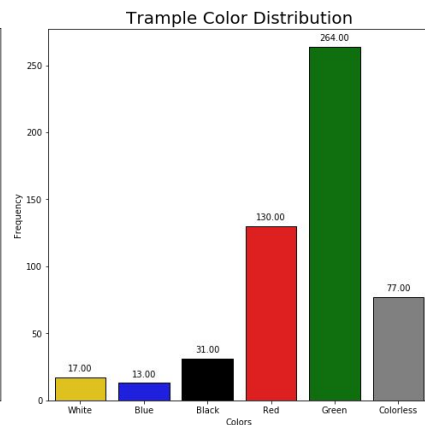
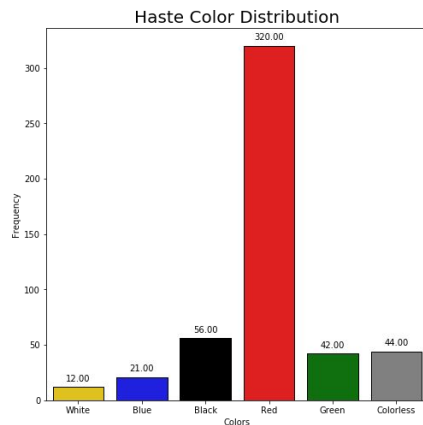
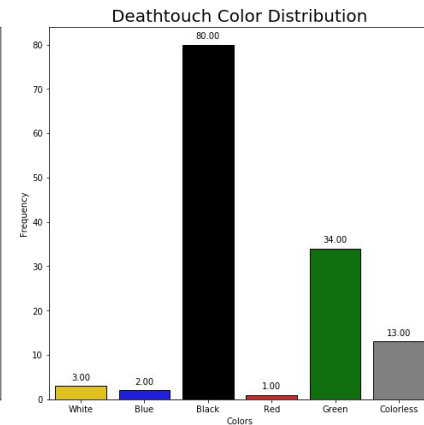
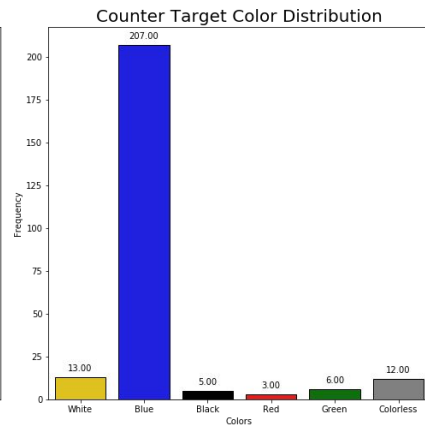
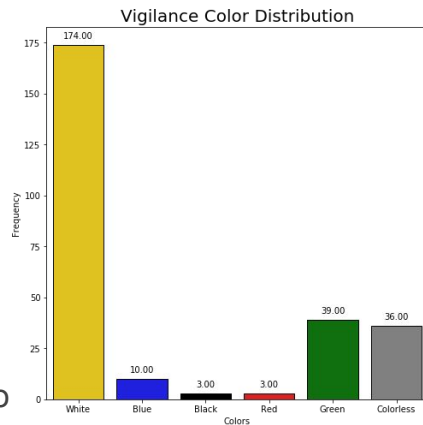
- White is more skewed towards lower mana costs while Green is more skewed towards higher mana costs.
- Colorless cards have the highest percentage of 0 cost cards.

Converted Mana Cost Distribution



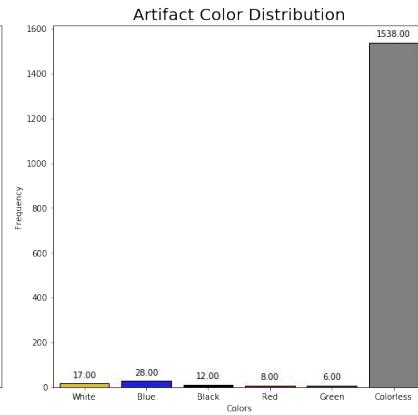
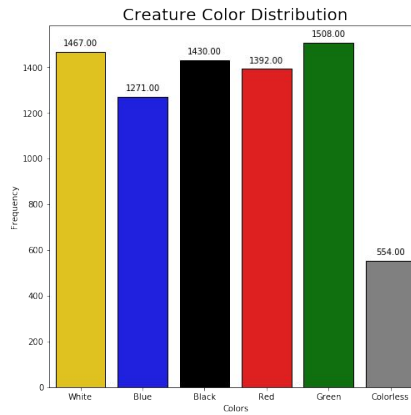
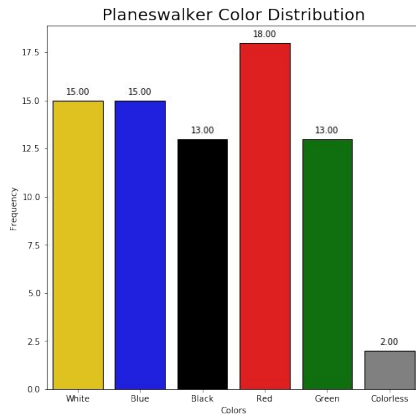
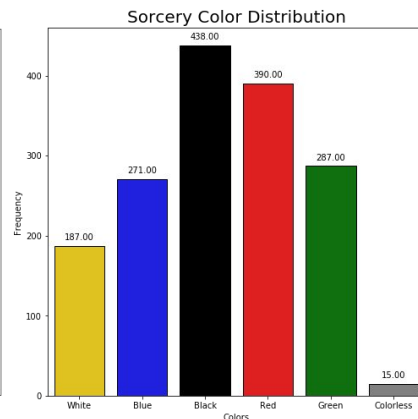
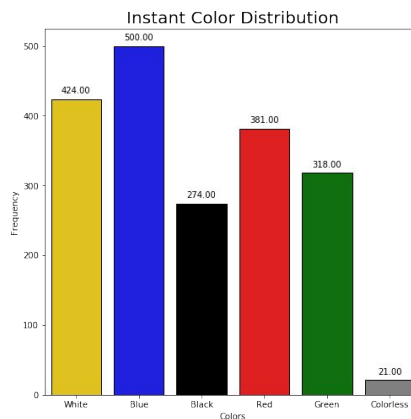
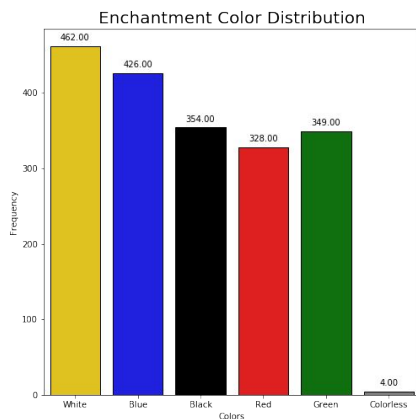
What makes the colors unique?

- Each color excels at a certain keywords.
- However, each color also has secondary and tertiary keywords that it has available to it.



What makes the color unique?

Each color also has a card type it likes to have.



What makes the color unique?

There are also features like:

- the ratio between the power and toughness for creatures
- the ratio between the mana cost and the stats of the cost for creatures.
- the ratio between the mana cost and the 'strength' of the effect.
- the number of times the card has been reprinted for standard.

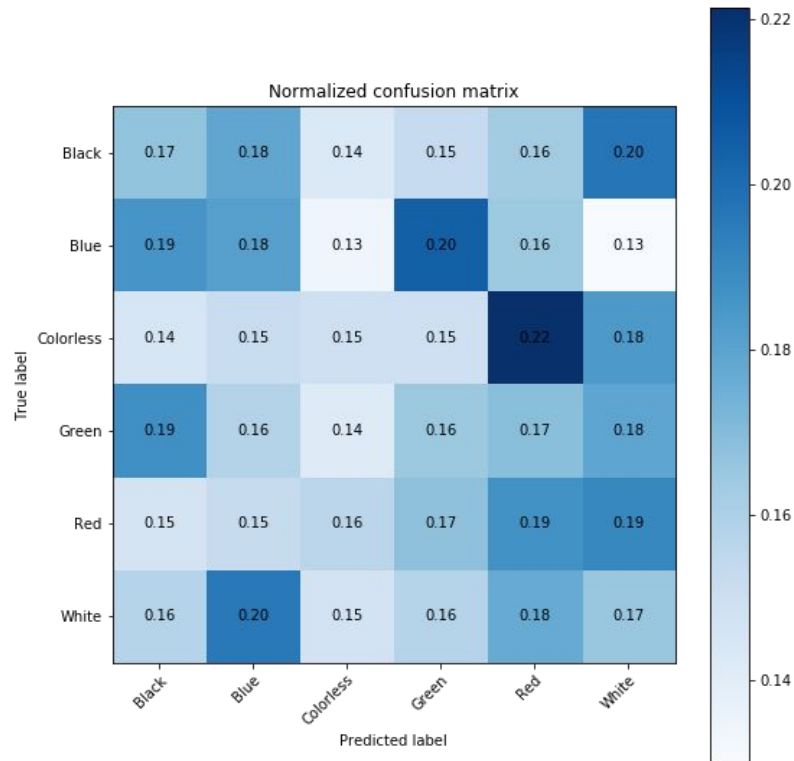


Don't be a Dummy

Ran a Dummy Classifier with the stratified strategy to serve as a baseline model.



Precision Score: 0.16884584345340858
Recall Score: 0.16898242064293326
Accuracy Score: 0.1692672998643148
F1 Score: 0.16883430443094496



More smart, less Dummy

Other models that I ran were:

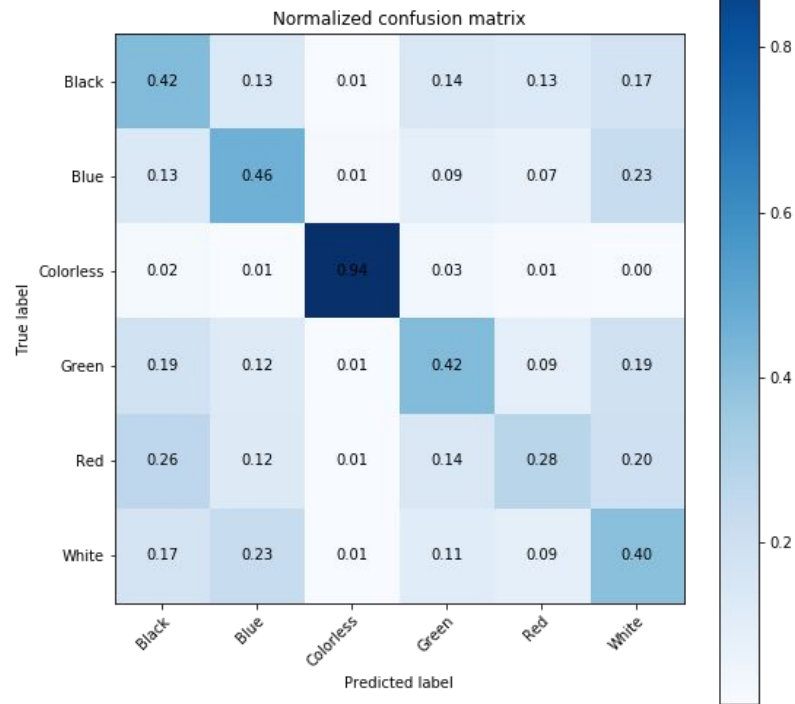
- KNN with tuned neighbors.
- Decision Trees with tuned max depth and max features.
- Bagging Trees and Random Forests.
- XGBoost tuned with GridSearchCV.



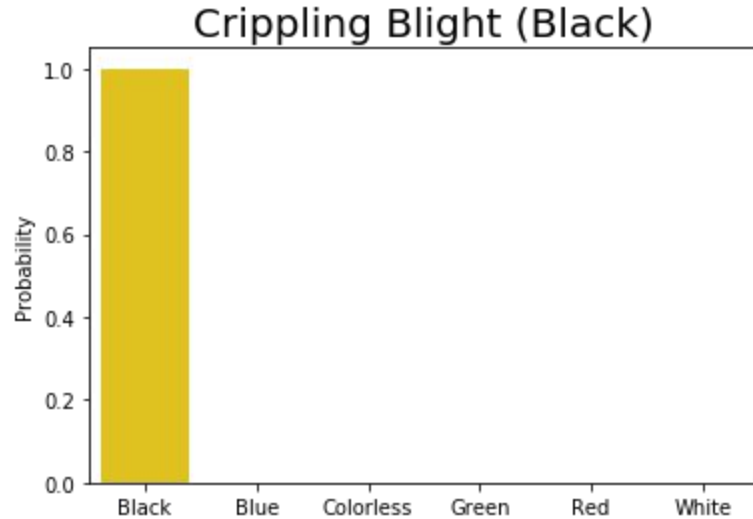
Final Model – XGBoost

- XGBoost was the most homoscedastic.
- Best metrics out of all the models.
- XGBoost was the best model after tuning with GridSearchCV.

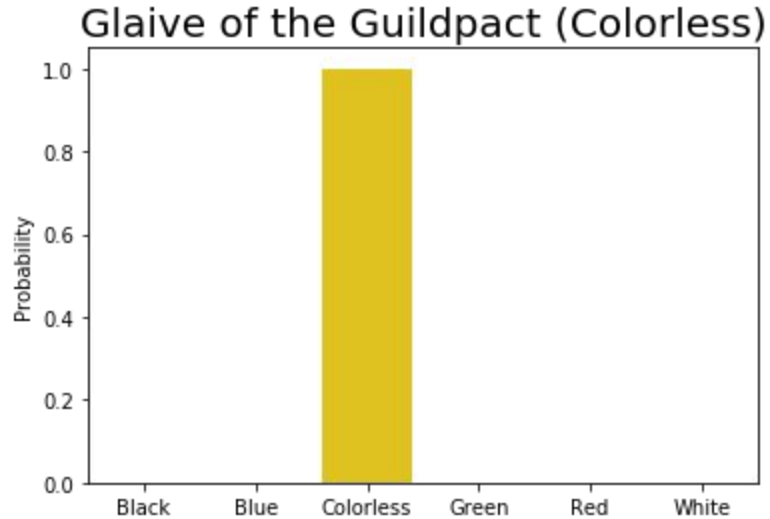
Precision Score: 0.4895217005826902
Recall Score: 0.4844306048445659
Accuracy Score: 0.4762550881953867
F1 Score: 0.48316299287347103



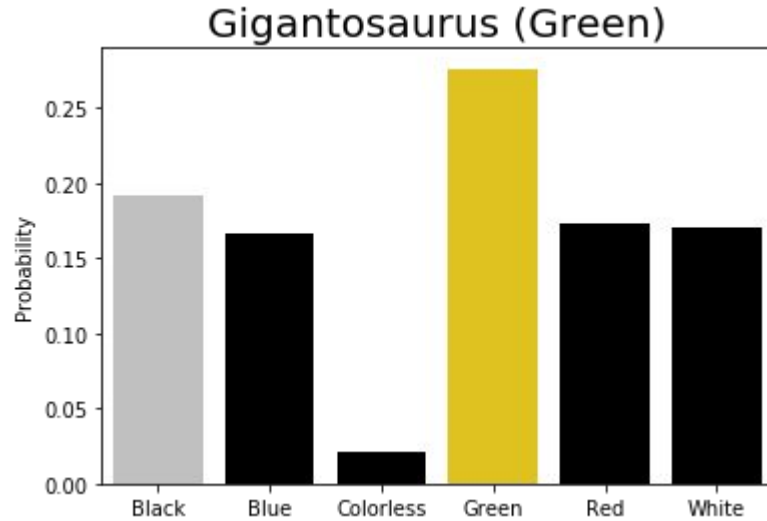
But does it work?



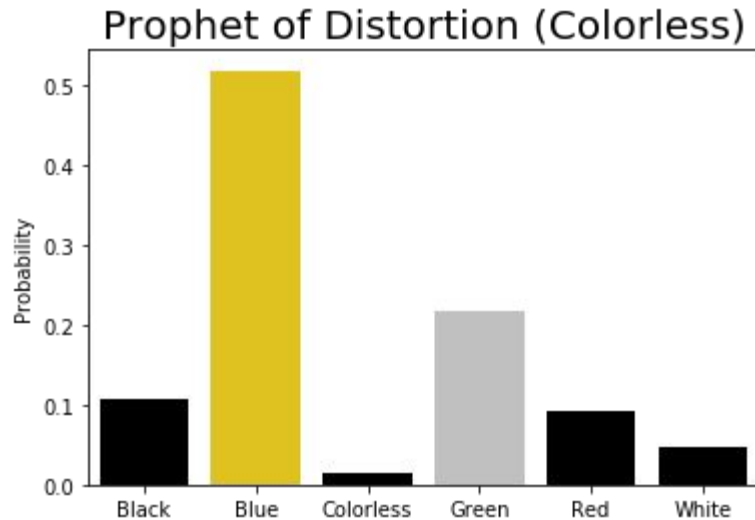
But does it work?



But does it work?

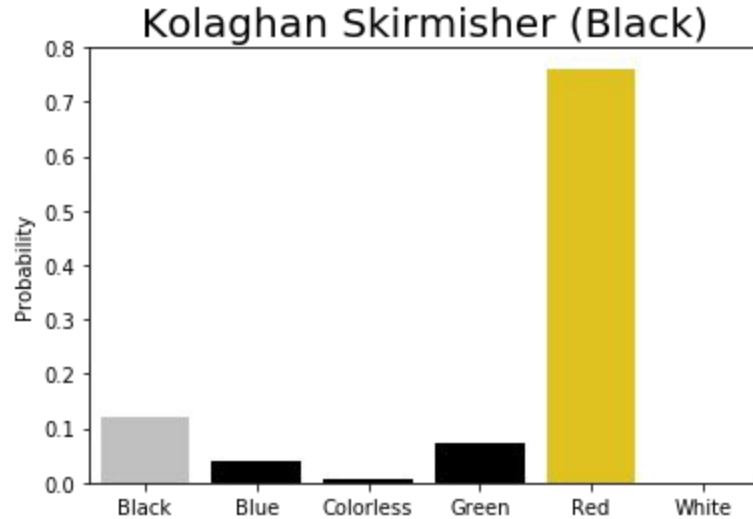


But does it work?

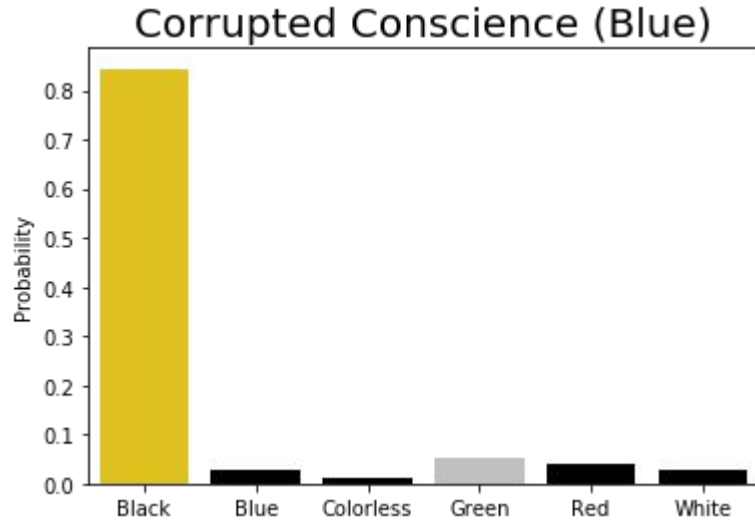


My model properly predicted the mana cost color of the card but I forgot to include 'devoid' as a keyword.

But does it work?



But does it work?



Model got tripped up on '-1/-1' in the infect description text.



Improvements

- Include more keywords/mechanics
 - Graveyard interaction for Black
 - Enemy creature control for Blue
 - Mana ramping for Green
 - Healing for White
 - Dealing direct damage for Red
- By better sorting out cards that belong to a color, the cards that don't fit any color identity can be confidently placed into the colorless category.
- Include a model that can predict if a card is multi-colored or not and then using the probability predictions to determining which color combination the card belongs to.

