

# Predictir

---

## Objective

- Create a system that takes existing card data (past prices, tournament usage, card attributes) and predicts future prices
- Be able to determine whether a card is worth purchasing or selling in a given time frame

## Motivation

- The market of the Magic the Gathering trading cards is a simpler and smaller version of the stock market
- Stock price prediction is a known and challenging problem: making progress on the simpler analogue could provide insight
- Players of the card game would like to know when it is best for them to buy and sell cards

# Big Card Price Data for

Boris A. Perkhounkov, Cooper Gates Frye, Et  
Stanford University CS229: Machine

## Data Extraction (cont.)

### Tournament Deck Data

- "Pro Tours":
  - Are highly competitive tournaments
  - Take place after new sets come out (4 times a year)
  - Are a deciding factor in the prices of new cards
- Data from past 12 (overlaps with price data time period)
- Collected top 8 deck lists for Pro Tours, counted appearances



Tournament 1, Winner's main deck	0
Tournament 1, Winner's side deck	0
...	
Tournament 5, Winner's side deck	0
Tournament 5, other main decks	5
Tournament 5, other side decks	0
Tournament 5, all decks	5
Tournament 6, Winner's main deck	0
Tournament 6, Winner's side deck	0
Tournament 6, other main decks	7
Tournament 6, other side decks	3
Tournament 6, all decks	10
Tournament 6, all decks	1
Tournament 6, all decks	0
Tournament 6, all decks	10
Tournament 6, all decks	0
Tournament 6, all decks	11
Tournament 6, all decks	0
...	
Tournament 12, all decks	0

# Appearances  
in Decks

- Want to know
- Use Support Vector Machine (SVM)
  - Unsure of linear models
  - SVR is a competitive model
- Preprocessing mean 0 and variance 1
- One input for ahead we are
- The price for tournament is a function of the price, so the price is a function of the price
- The price in tournament is a function of the price, so the price is a function of the price
- Each run split 70% training and 30% testing
- We choose the model with the lowest error

### Set Membership Data

- Affects supply (older sets are not printed)

# Magic the Gathering

nily Margaret Franklin

Learning

## Methodology

future price: a regression problem.

Vector Regression

arity of the problem

putationally efficient method,

: standardized features to have  
variance 1

our system is the number of days  
predicting

tomorrow will usually be close to today's

prediction will appear to be too accurate.

70 years is affected by many random

is hard to make use of accurate predictions.

is the data into 70% training and

oints

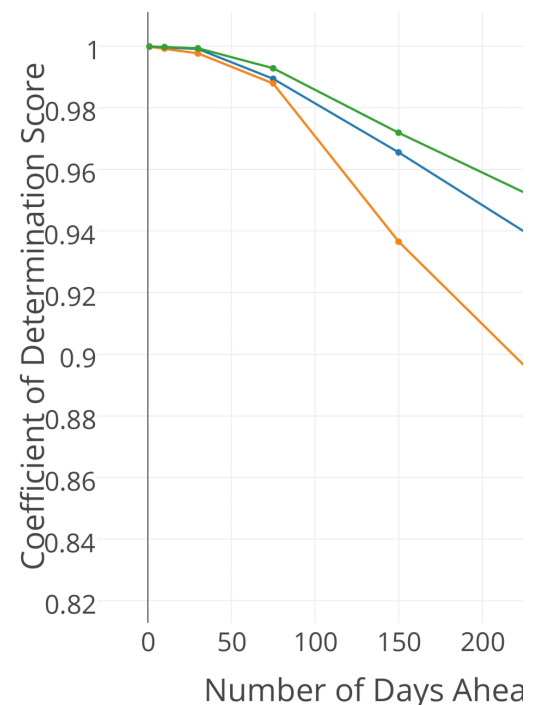
e hyperparameters by grid search,

th 7-fold cross validation.

## Results

## Result

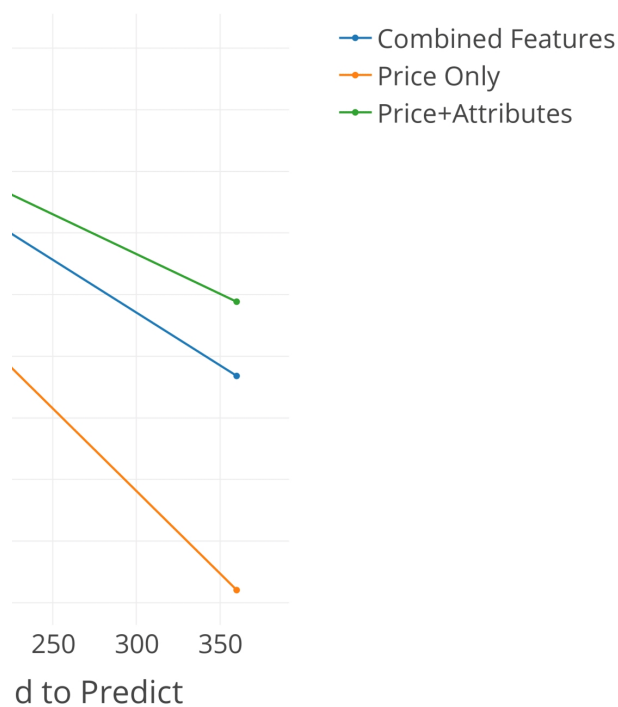
Varying How Far In



- As expected, accuracy decreases as the prediction horizon increases.
- The methods which rely on historical prices are more affected by noise.
- Results in this plot are for 1500

## ts (cont.)

into the Future to Predict



$r$  for all methods decreases  
prediction time span

ely more heavily on past  
ed.

e from a training set of size



# Data Extraction

- 15440 distinct cards in the game
- Historical Price Data
  - Scraped from <http://www.mtgstocks.com/>
  - 26348 cards' data (repetitions from reprints)
  - Price data ranges from June 9, 2012 - Nov. 6, 2015 (filled in with 0s)
  - Ignored cards that cost less than \$1 (80%) - they aren't worth it to buy/sell because of transaction costs, and they have very low variance in price, so they aren't very predictive



Snapcaster Mage (R)

Innistrad

\$51.01 \$59.89 \$99.99

Finance

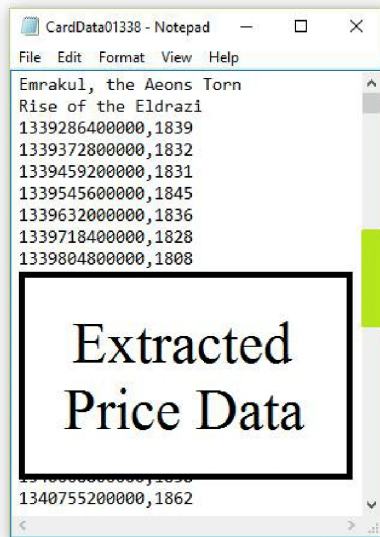
Card



Price on 2012-06-09 17:00:00 2133  
Price on 2012-06-10 17:00:00 2146  
Price on 2012-06-11 17:00:00 2111  
Price on 2012-06-12 17:00:00 2097  
Price on 2012-06-13 17:00:00 2083  
...  
Price on 2015-06-08 17:00:00 7390  
Price on 2015-06-09 17:00:00 8000  
Price on 2015-06-10 17:00:00 8424  
Price on 2015-06-11 17:00:00 8424  
Price on 2015-06-12 17:00:00 8399  
Price on 2015-06-13 17:00:00 8525  
Price on 2015-06-14 17:00:00 8550  
Price on 2015-06-15 17:00:00 8504  
Price on 2015-06-16 17:00:00 8550  
Price on 2015-06-17 17:00:00 8599  
Price on 2015-06-18 17:00:00 8549  
Price on 2015-06-19 17:00:00 8499  
Price on 2015-06-20 17:00:00 8424  
Price on 2015-06-21 17:00:00 8343  
...  
Price on 2015-11-04 16:00:00 6109  
Price on 2015-11-05 16:00:00 6172  
Price on 2015-11-06 16:00:00 6068

Price Data

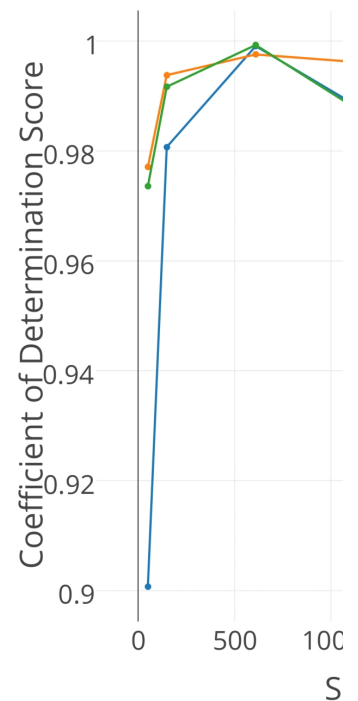
- Tournament legality is partially determined by set



10th Edition	0
7th Edition	0
8th Edition	0
9th Edition	0
Alara Reborn	0
Alliances	0
Alpha Edition	0
...	
Innistrad	0
Invasion	0
Journey into Nyx	1
Judge Promos	0
Judges	0
Kharidian	0
...	
Worldwake	0
Zenith	0
Zendikar	0

**Set**

- All prediction
- Tried  $C = 1, 2$
- Linear kernel polynomial ke



## Card Attributes Data

- Downloaded JSON files containing card information from mtgjson.com
- Parsed by in-game cost, keywords in description of usage
- Such attributes determine usefulness and play a large factor in demand



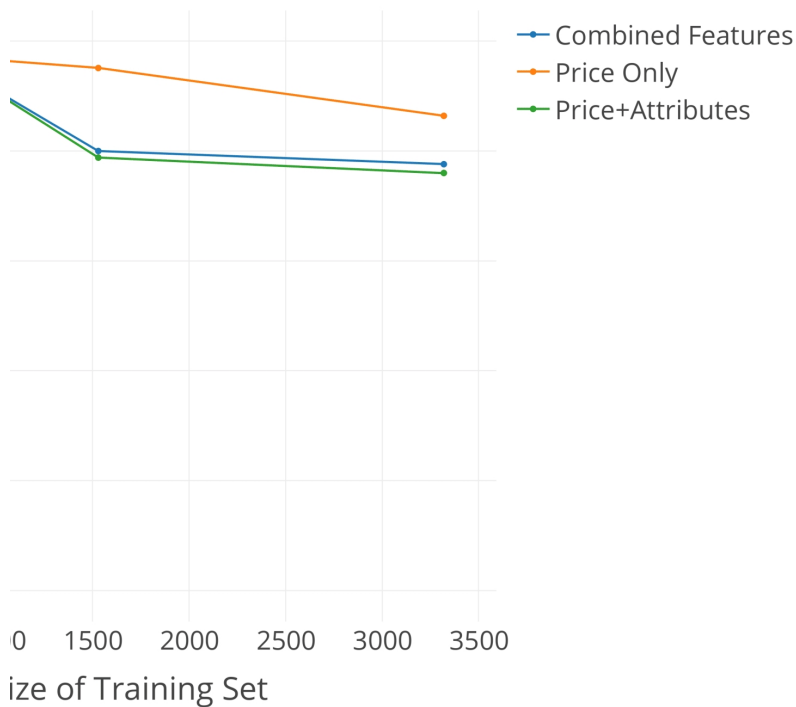
Generic Mana	2
{W}	0
{U}	1
{B}	0
...	
Converted Mana Cost	3
White	0
Blue	1
Black	0
Red	0
...	
Modern	1
Legacy	1
Vintage	1

**Card Attributes**

- Accuracy dec
- Cross validation alleviate this s

s 30 days forward  
 10, 100, 500 with  $C = 100$  the best  
 performed better than radial or  
 kernels

Varying Size of Training Set



ays unexpectedly  
 on with random sampling will help  
 artifact

## Analysis and

- Most surprising result from the tournament
- We suspect there are at fault
  - The number of cards (es tournament play is very
  - There are more tournam demand from what our c
- Decks beyond the top could help, but the d
- The other surprising accuracy was.
- May be a result of fin cards, where predictio
- Predicting those has l make a profit by buyi
- Next step is to redefin tracks cards with dyn
- Then we will use that further modify our fea

# 1 Further Work

---

was the lack of impact data.

two main effects currently

p. expensive cards) that see small.

ents played that gradually change data currently captures.

8 and smaller tournaments data is unavailable

result is how high our

ancially stable expensive n is a trivial task

ow utility, since you can't ng them or selling them.

ie accuracy so that it only amic prices

new accuracy metric to ture selection.