Spotting Cinderella at the Dance: Predicting Upsets in the NCAA Men's Basketball Tournament

Matthew Worley

Presentation Outline

Introduce and define the problem

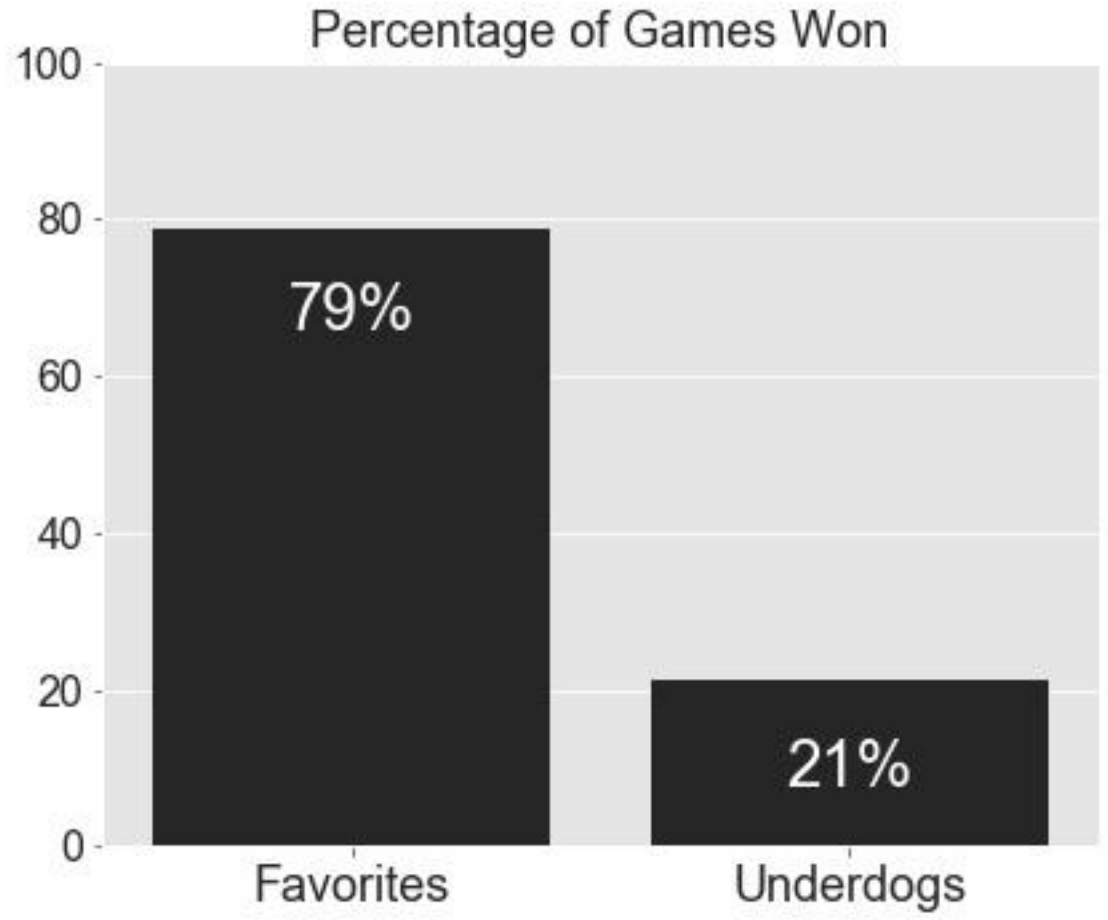
Performance: Metrics and Real-World

Adding value and iterating to improve

The Problem

Most games are an uneven matchup between a favored team expected to win and an “underdog”

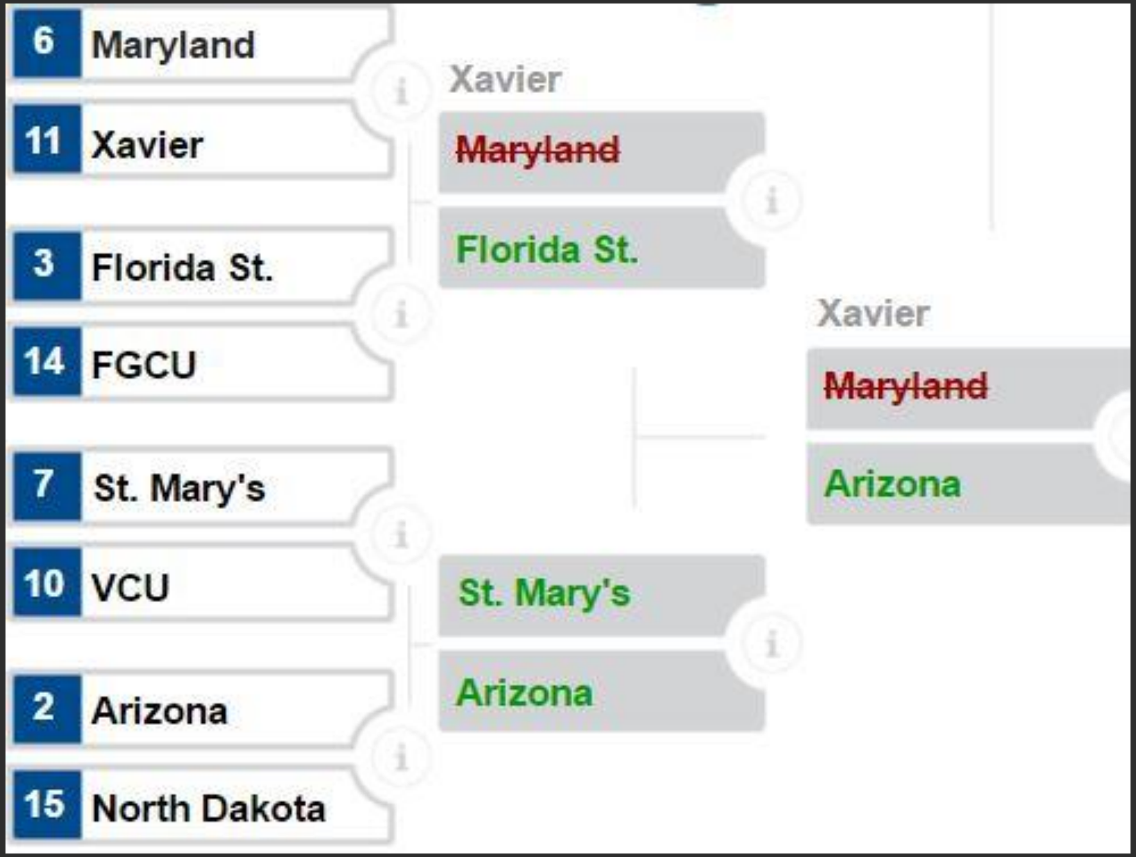
Around one-fifth of the time, underdogs “upset” the favorites



Upsets bust brackets

Upsets cause bracket pool players to lose points

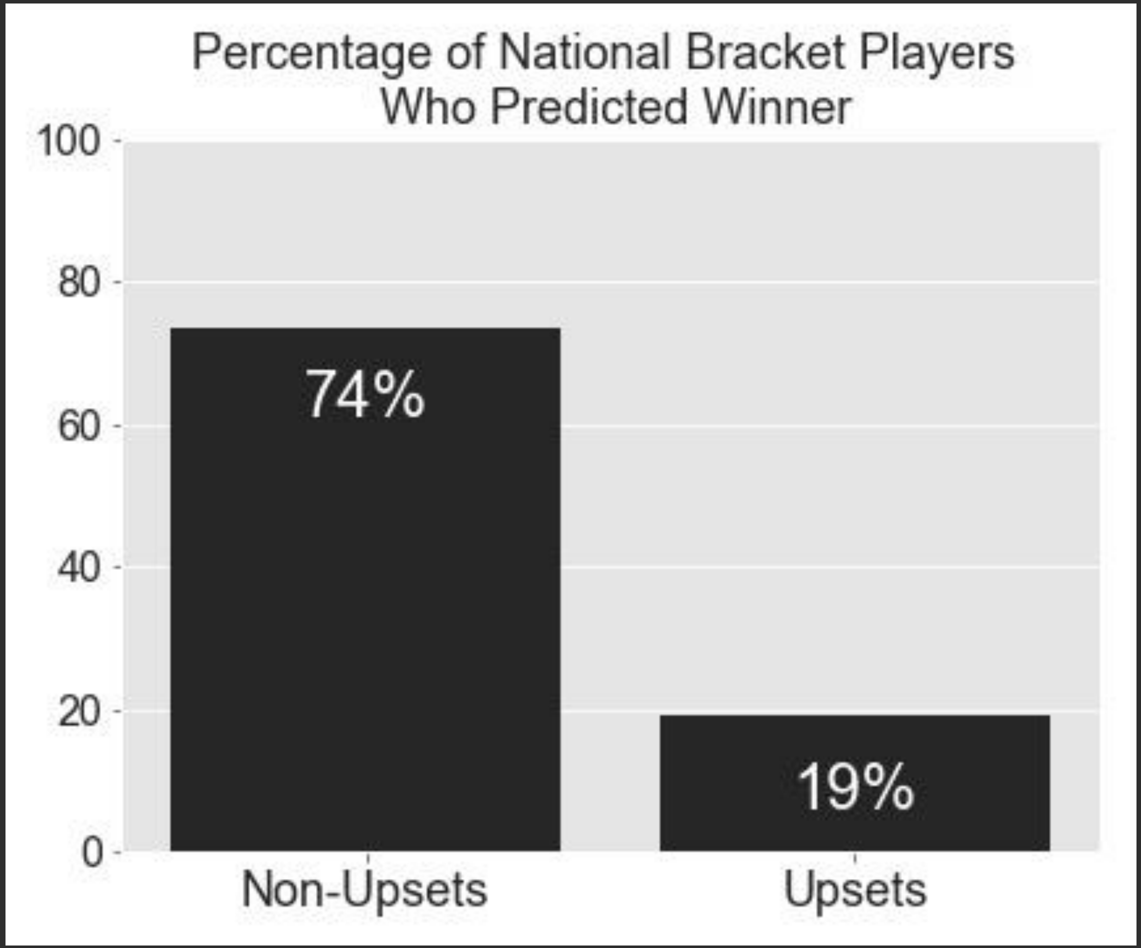
Future matchups are eliminated



Which upsets to predict?

Bracket players have worse accuracy for upsets

Know they occur, but lack guidance on which to pick



“Experts” or Algorithms?

The Solution

Use machine learning to provide better upset guidance for bracket contests and game wagers

Project Pipeline

Data Processing and Exploration → **Create and Compare Models** → **Test Models**

Data Processing and Exploration

- Download and web-scrape data
- Build characteristics for each tournament team
- Label games as upsets or not

Create and Compare Models

- Train classification models to predict upsets from features of teams in each game

Test Models

- Use best models to classify held-out games
- Provide estimate of future performance

Strategy: Broad range of team features

Player Statistics

GAME STATISTICS

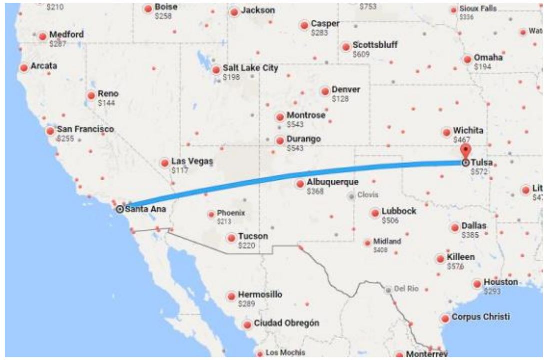
Player	GP	MIN	PPG	RPG
Frank Mason III	36	36.1	20.9	4.2
Josh Jackson	35	30.8	16.3	7.4
Devonte' Graham	36	35.3	13.4	3.1
Sviatoslav Mykhailiuk	36	27.3	9.8	3.0
Landen Lucas	35	25.6	8.0	8.3

GAME STATISTICS

Player	GP	MIN	PPG	RPG
Brynton Lemar	36	33.0	16.1	3.3
Chima Moneke	36	27.3	14.6	9.5
Siler Schneider	36	23.2	10.3	3.3
Lawrence White	36	27.3	7.7	3.6
Darius Graham	36	28.9	7.3	1.8

Coach History

Distance Traveled



Team Season Performance

Rank	Team	Conf	W-L	AdjEM
6	Kansas 1	B12	31-5	+27.45

Rank	Team	Conf	W-L	AdjEM
218	UC Davis 16	BW	23-13	-4.41

1 Kansas (28-4)

16 UC Davis (23-12)

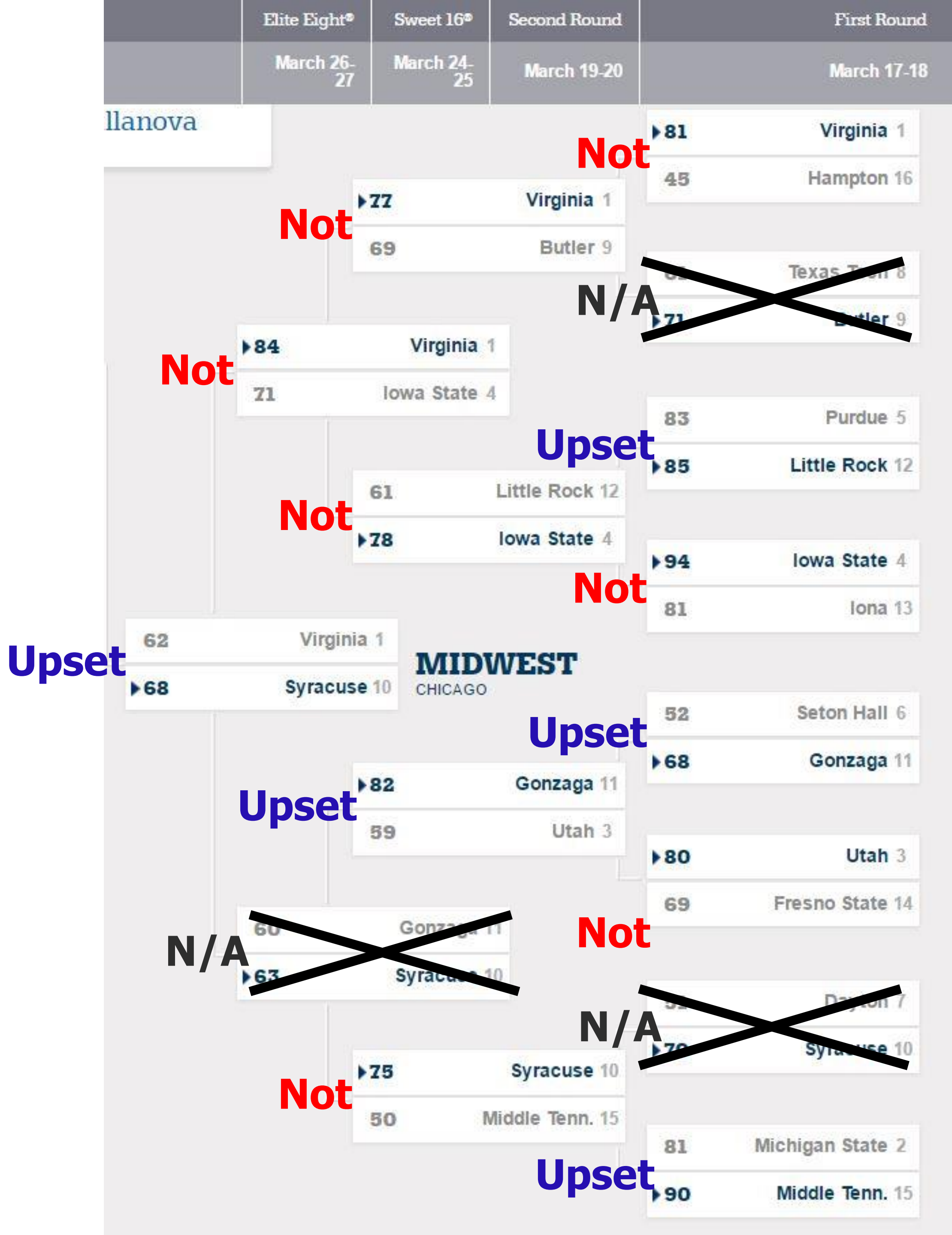
Focus only on upsets

Games with "upset potential" involve a "favorite" vs. an "underdog"

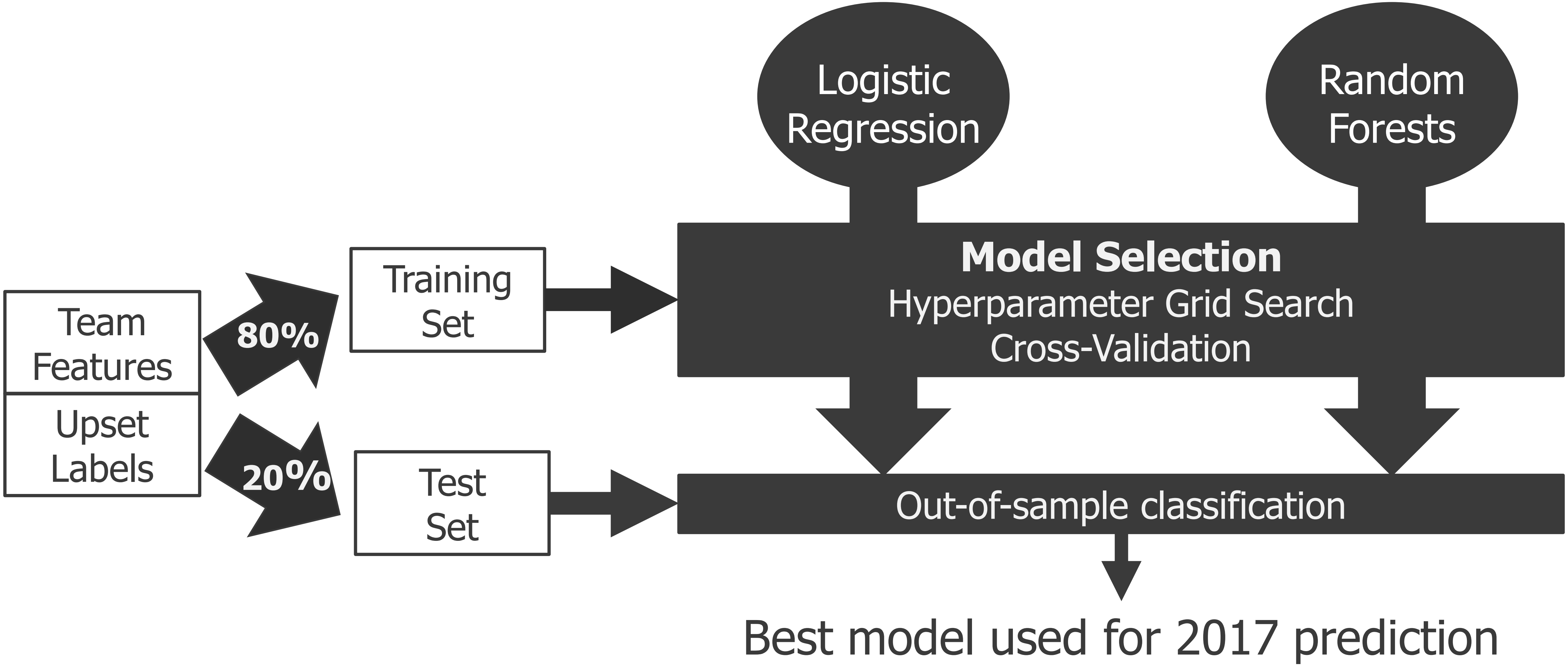
Criteria: Seed difference > 3

Labels for games that meet criteria:

"Upset" - Underdog wins
"Not upset" - Underdog loses
"N/A" – Game not used



Machine Learning Process



Use Case: 2017 NCAA Tournament

Use model to classify future games

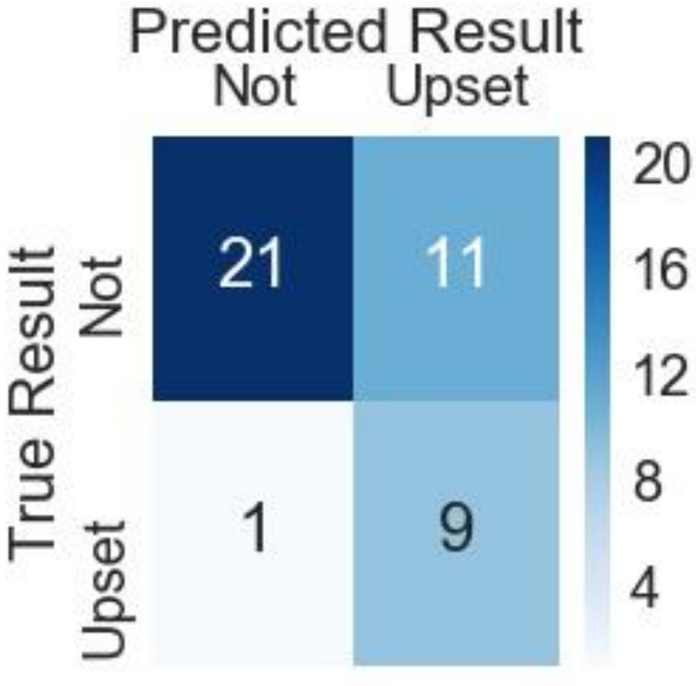
90% Upset Recall

Predicted 9 of 10 true upsets

45% Upset Precision

9 true upsets in 20 upset predictions

$F1 = 0.60$

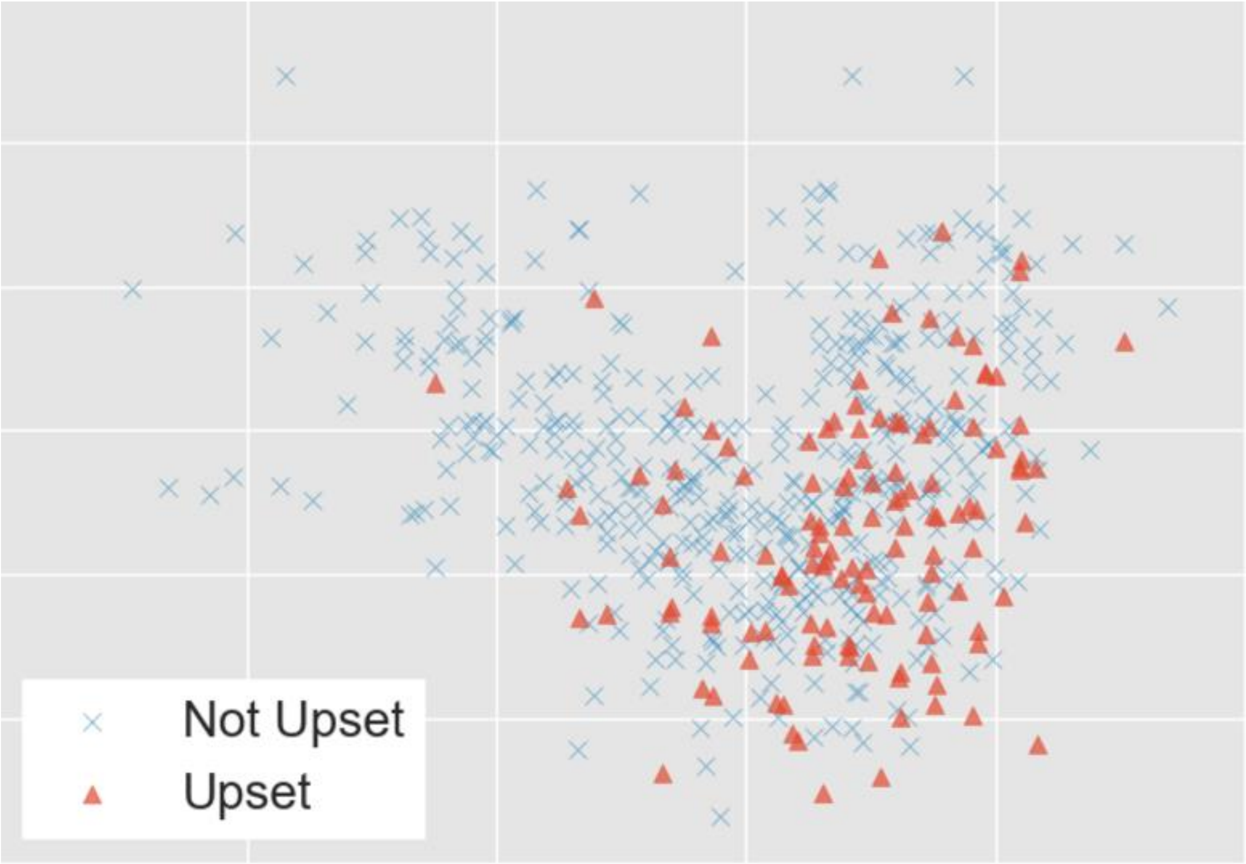


17% return on wagers using predictions

Scenario: Place $100 money line bet on each game

30 Correct Predictions	$1900	Net Profit	$700
12 Incorrect Predictions	- $1200	Total Investment	÷ $4200
Net Profit	$700	ROI	17%

Iteration: Addressing Label Imbalance



Only ~20% of games are an upset

Most classification algorithms perform better with more balanced labels

The Solution

Use "resampling" to reduce the label imbalance in training sets

Random Undersampling

Randomly remove non-upsets

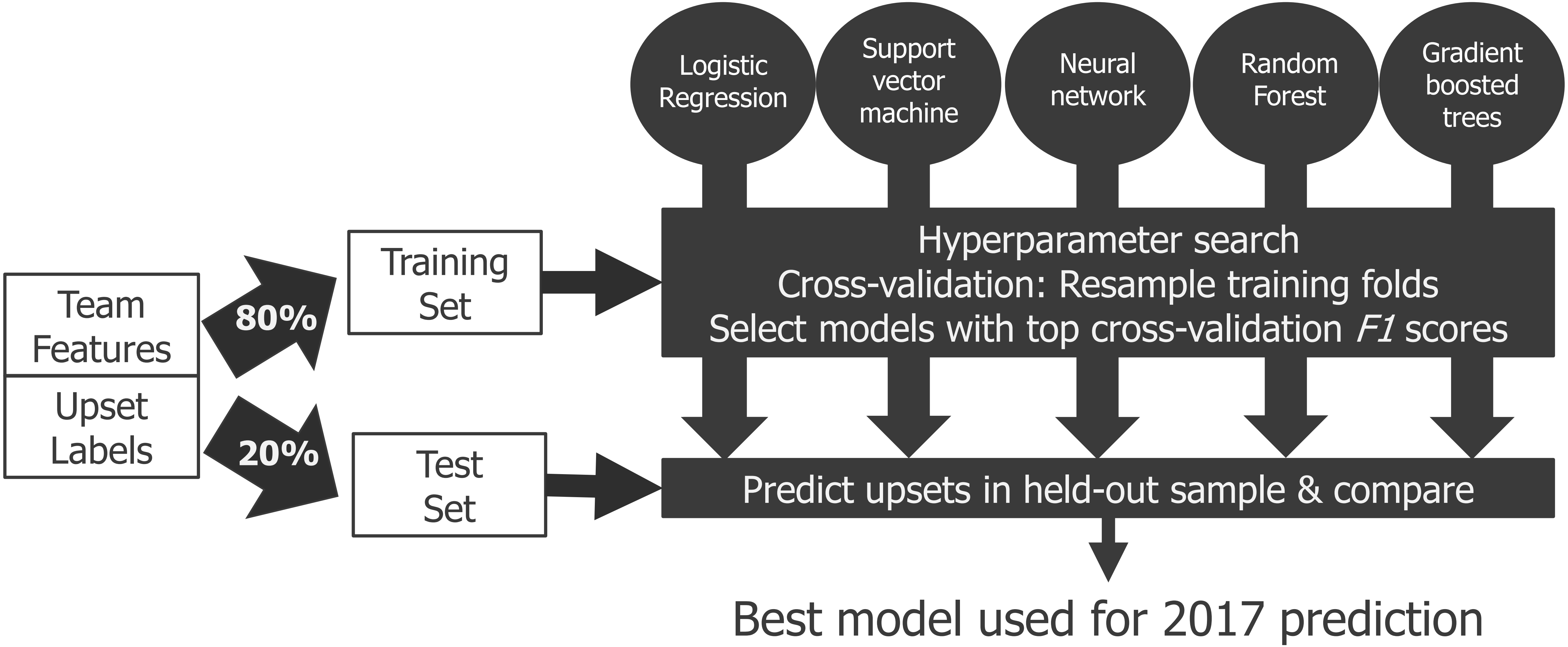
Random Oversampling

Randomly repeat upsets

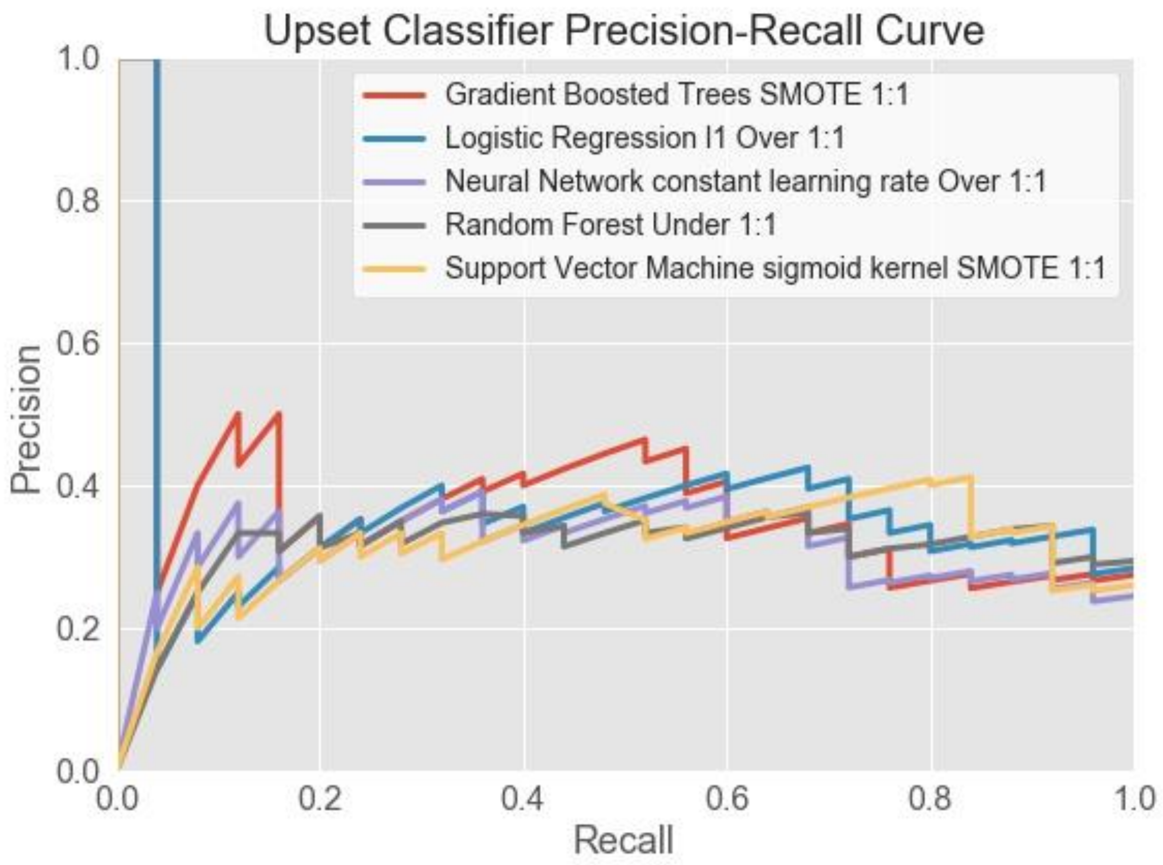
Synthetic Oversampling

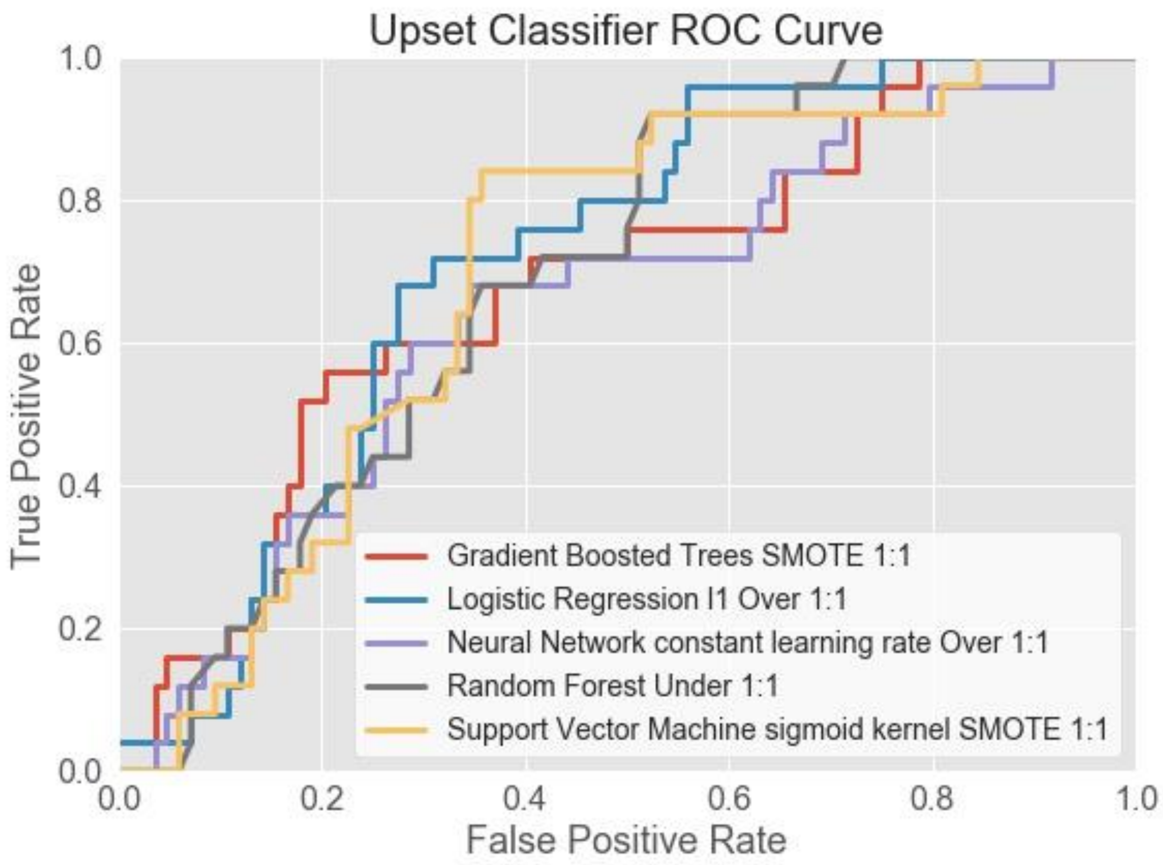
Create new upsets similar to authentic ones

Resampling in Machine Learning



Model Selection: Test set performance





Use Case: 2017 NCAA Tournament

Resampling improved upset classification

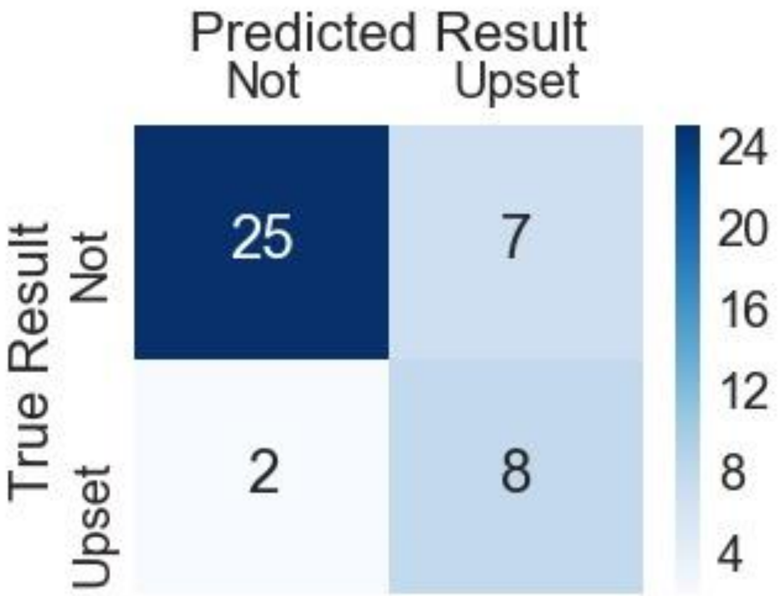
80% Upset Recall

Predicted 8 of 10 true upsets

53% Upset Precision

8 true upsets in 15 upset predictions

F1 = 0.64

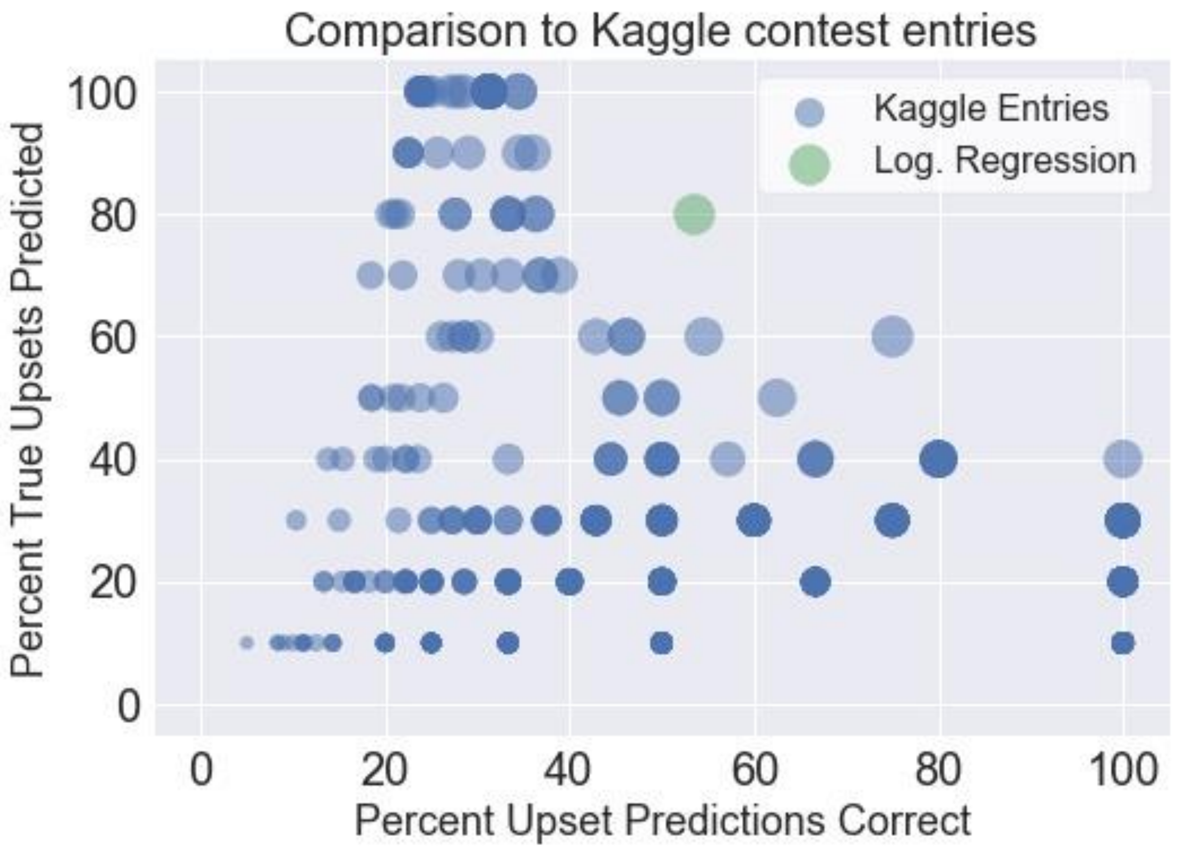


Scoring the model vs. Kaggle submissions

A trade-off between predicting and few false positives (x-axis) and most true upsets (y-axis)

Model achieved good balance

In 1st percentile of Kaggle contest entries (one entry had higher *F1*)



21% return on wagers using predictions

Scenario: Place $100 money line bet on each game

33 Correct Predictions	$1789
9 Incorrect Predictions	- $900
Net Profit	$889

Net Profit	$889
Total Investment	÷ $4200
ROI	21%

Client Recommendations

Bracket Contest Players

Pick Round 1 underdog winners

79% accuracy rate in 2017

Differentiate scores from competitors

Be skeptical of favorites predicted to be upset

None made Final Four in 2017

Game Wagers

Weight bets to high-confidence picks

Moderate-return underdogs

Small-return favorites

Spread smaller bets across lower-confidence upset picks

Account for most false positives

Pick games with larger returns

Lessons Learned

Tailor Scoring and Tuning for Project Purpose

For upsets, accuracy not helpful

Metric depends on the use case and client goal (cost/benefit)

Iterate and optimize at any point in the chain

Feature engineering > feature selection > model tuning > model selection

Experimentation not deliberation

Enthusiasm and Energy

Work on problems that interest you

Product or experience may have unforeseen applications

2018 Tournament

True		Predicted: Not	Predicted: Upset
	Upset	**22**	**5**
	Not	**5**	**5**

Precision	.50
Recall	.50
F1	.50
Accuracy	.73

FN: UVA, Wichita St., Arizona, UNC, Loyola-Chicago vs UT

FP: SDSU, New Mexico St., Butler, Rhode Island, Seton Hall