

# Wages of Wins

Albrecht Zimmermann, Université de Caen

19/09/2016, MLSA '16 @ECML/PKDD

# Motivation

- Previous work : predicting match outcomes
- Decent accuracies but what for ?
- MLSA '13 recommendation : beat the bookie !
- Could one actually make money ?

# Setting

- Predictions : NCAAB « march madness », NBA, NFL – all 2015/2016
  - Outcomes, not scores Haven't managed yet to train good score predictors
  - Basketball as in earlier work, Football described later
- Data: [www.kenpom.com](http://www.kenpom.com),  
[www.basketball-reference.com](http://www.basketball-reference.com),  
[www.pro-football-reference.com](http://www.pro-football-reference.com)
- Money Lines : [www.vegasinsider.com](http://www.vegasinsider.com)

# Money Lines

- Goal for sports book : make money !
  - Partition betting volume s.t. bettors' wins (+profit) covered by losses
  - Adjust « money lines »

# Money Lines (2)

	Match-Up	Favorite	Underdog	Fav-Line	DogLine
1	Detroit at Atlanta	Atlanta	Detroit	300	240
2	Utah at Detroit	Detroit	Utah	110	-110

- Match 1 :

- Correct bet Detroit, 100\$ bet = 240\$ gain
- Correct bet Atlanta, 300\$ needed to win 100\$ → 100\$ bet = 33.33\$ gain

**Atlanta clear favorite :  
motivate people to bet  
on Detroit**

- Match 2 : « Pick 'em »

- Correct bet of 100\$ = 90.90\$ gain

**Remember : wins + PROFIT**

- Initially based on model, bettor behavior shifts lines

**Correct Dog-Bet > Correct Pick 'em-Bet > Correct Fav-Bet**

# Simulated betting

- Bet all matches
  - 100\$ per match
  - Separate regular/post-season for NBA/NFL
- At money-line w/smallest spread
  - E.g. 170/150 instead of 200/170

Conservative but NOT as conservative as I thought !

- Tally winnings per day
  - And over full period

# « Vegas accuracy »

- « Predicts » only favorites
- Pick 'ems → Coin flips →  
50 % correct expected
  - Best case all correct
  - Worst case all wrong

Can make big  
difference financially

# NCAAB (67 matches)

## Vegas

5 Pick 'ems

w/o Pick 'ems		w/ Pick 'ems					
Accuracy	Pay-out	Exp. Acc	Pay-out	Best Acc.	Pay-out	Worst	Pay-out
0.6612	30.26	0.6417	7.51	0.6865	484.76	0.5821	-469.73

41 Favs

## Predictors

Classifier	Naïve Bayes	Multilayer Perceptron	Simplified KenPom
Accuracy	0.6865	0.6417	0.7014
Pay-out	-30.83	-834.81	-231.34

## Distribution of predictions

Classifier	Favs	Dogs	PEs
NB	42	2	2
MLP	40	0	3
KP	43	0	4



# NCAAB (67 matches)

5 Pick 'ems

w/o

Accura

0.6612

41 Fa

-out

9.73

Class

Accura

Pay-ou

Pom

**Something's not  
right here !**

RI

43

0

4

# NBA (1288, 115 PEs)

## Vegas

w/o Pick 'ems		w/ Pick 'ems					
Accuracy	Pay-out	Exp. Acc	Pay-out	Best Acc.	Pay-out	Worst.	Pay-out
0.7121	-2374.16	0.6865	-1857.3	0.7375	9125.84	0.6492	-12828.81

835 Favs

## Predictors

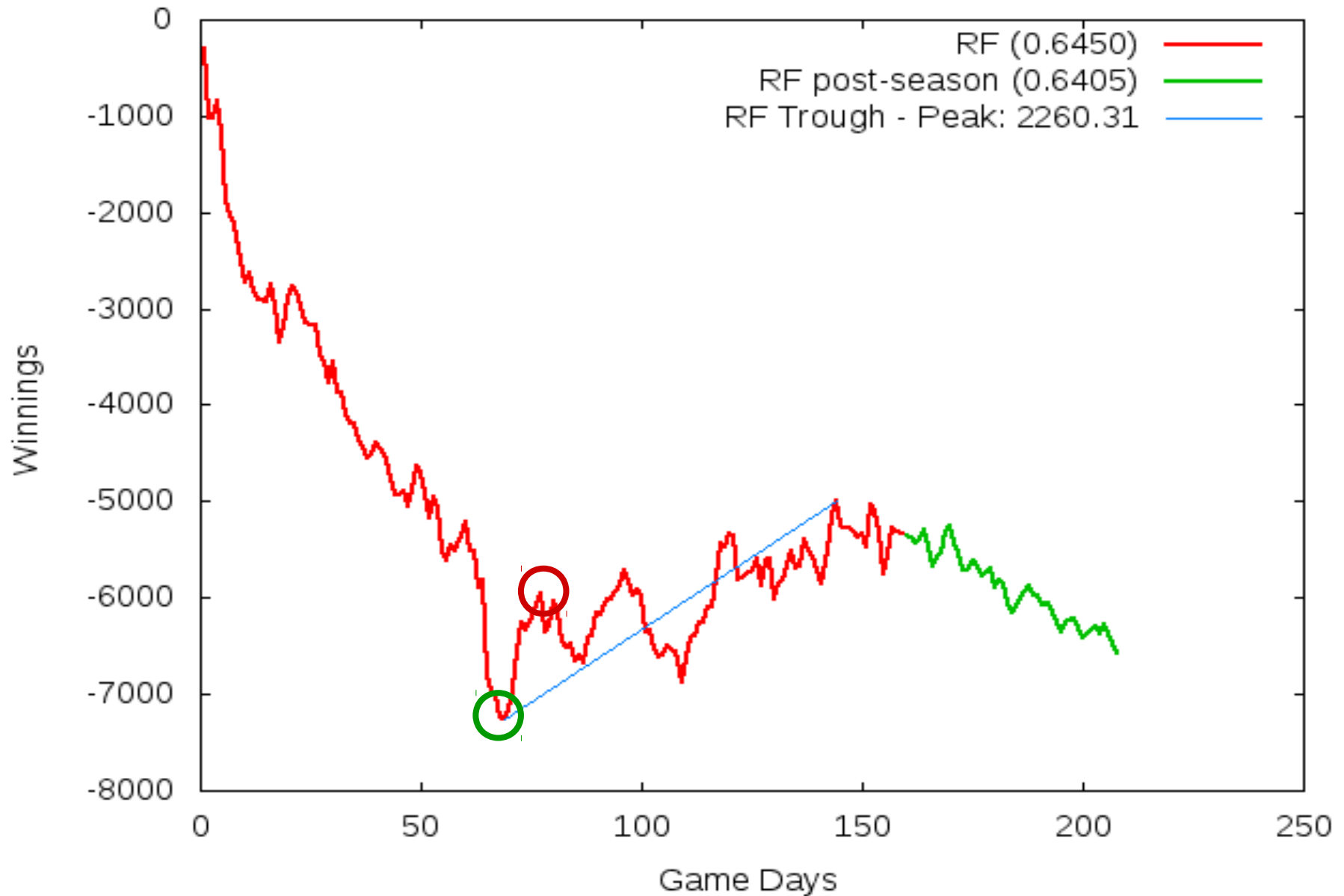
Classifier	NB	MLP	Random Forest	KP
Accuracy	0.6607	0.6615	0.6405	0.6669
Pay-out	-2882.21	-2005.56	-6580.88	-1543.05

## Distribution of predictions

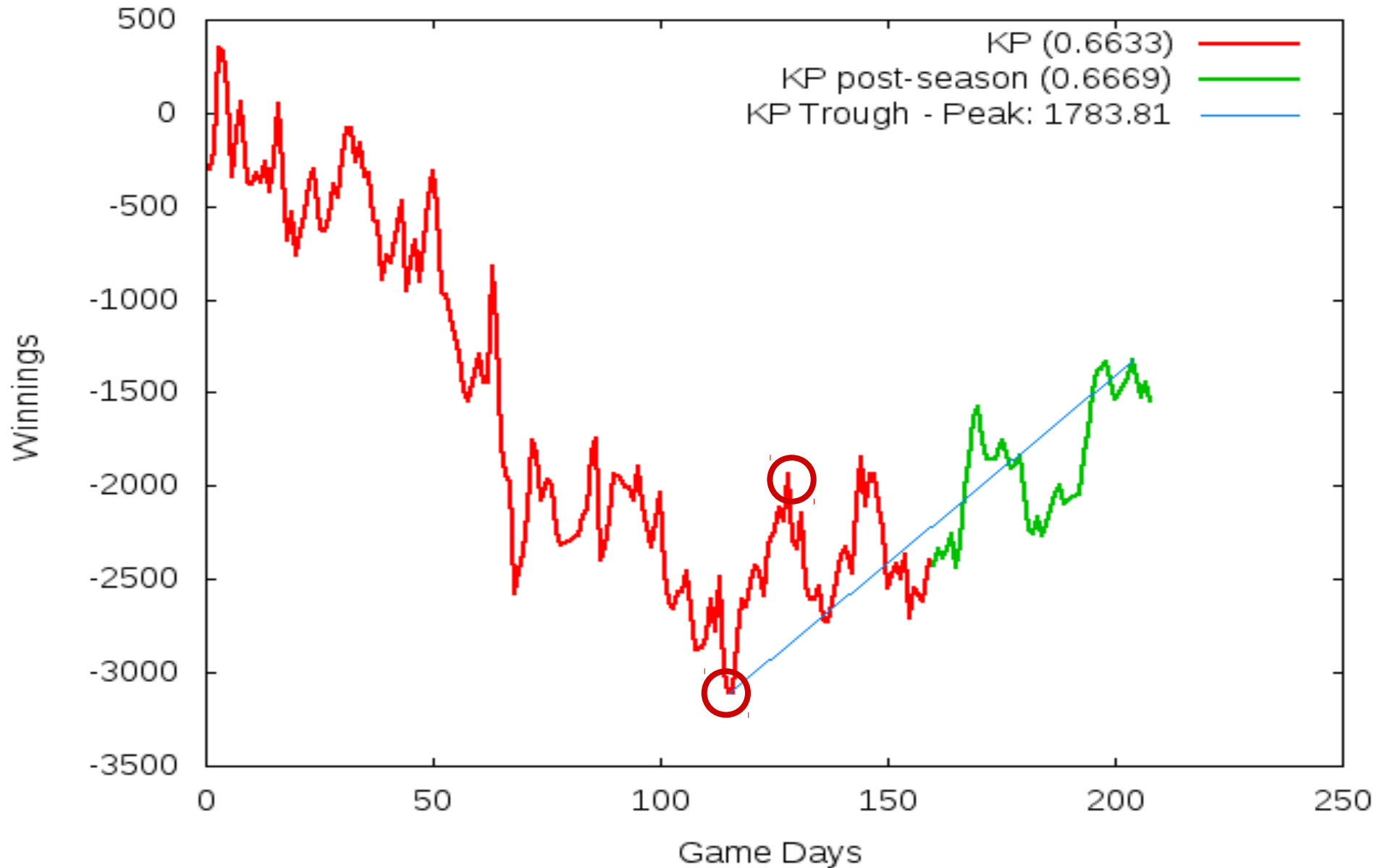
Regular +  
post-season

Classifier	Favs	Dogs	PEs
NB	691	57	48
MLP	707	60	22
RF	685	61	28
KP	725	59	12

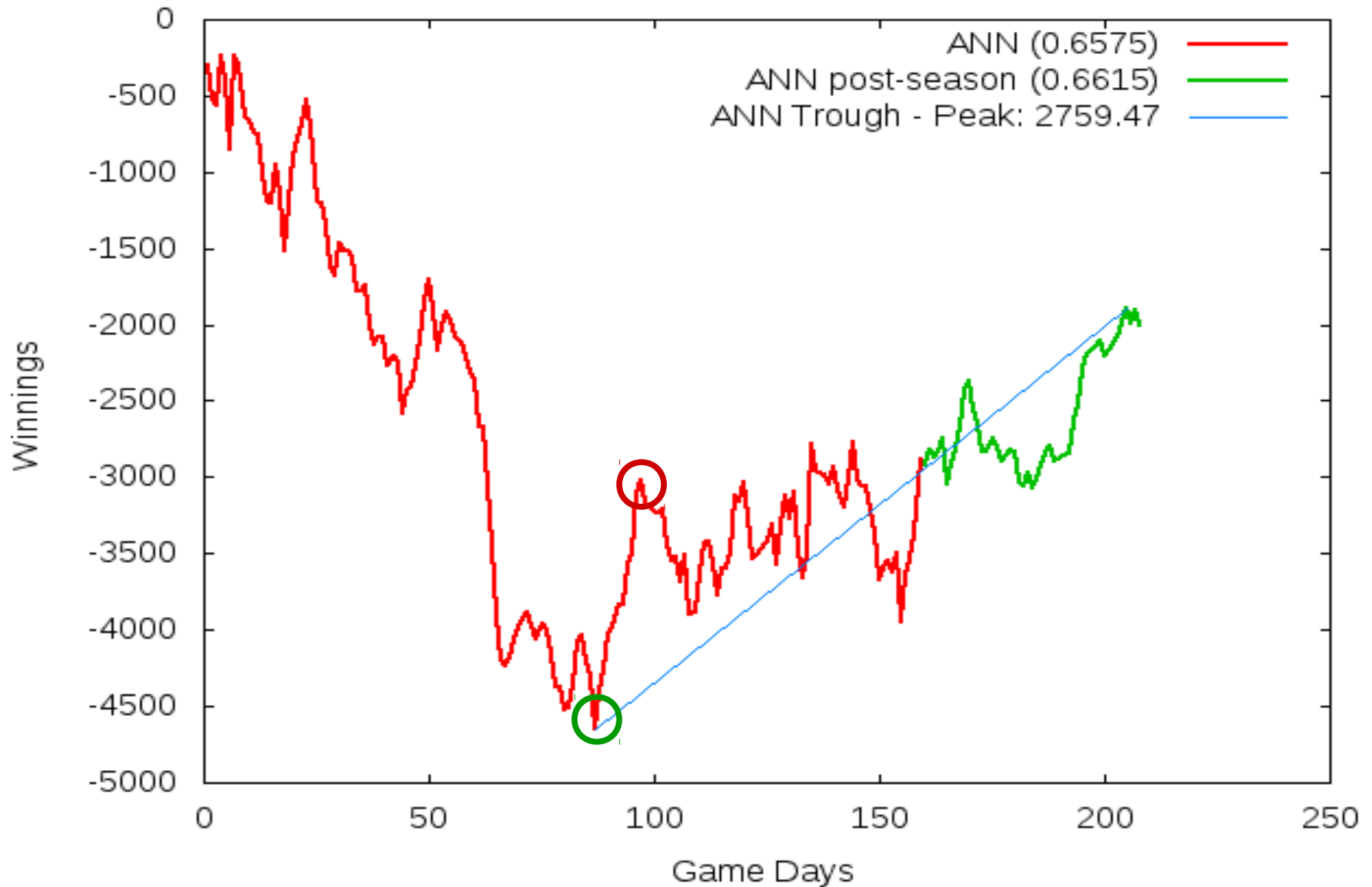
# RF winnings curve



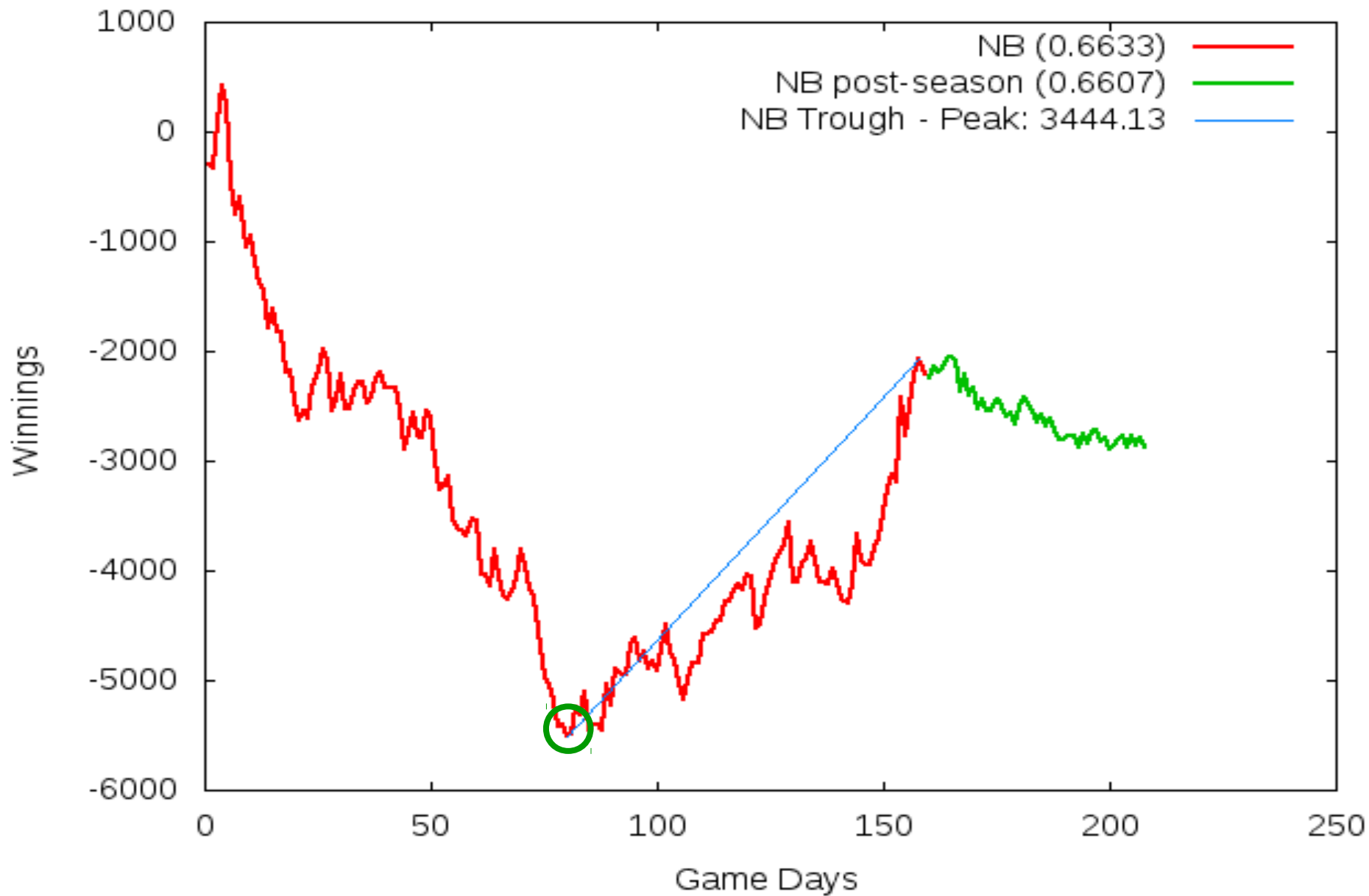
# KP winning curve



# MLP winning curve



# NB winning curve



# NFL prediction

- Different game :
  - Different outcomes for possessions
  - Not many scoring events
  - Two (three) distinct sub-teams
  - Fewer matches
- State of the art : « Defense-adjusted Yards above Replacement » (DYAR)
  - Needs fine-grained data, expensive calculation
- My « solutions » :
  - Averaged (adjusted) statistics
  - Opponents' statistics
  - Strength of schedule

# NFL (251, 29 PEs)

## Vegas

w/o Pick 'ems		w/ Pick 'ems					
Accuracy	Pay-out	Exp. Acc	Pay-out	Best Acc.	Pay-out	Worst	Pay-out
0.6441	-1215.69	0.6294	-1251.92	0.6852	1420.68	0.5697	-4115.42

143 Favs

## Predictors

Classifier	NB	MLP	RF	Simple Rating System
Accuracy	0.6335	0.5896	0.5737	0.5896
Pay-out	1777.98	729.92	-1591.68	-1255.71

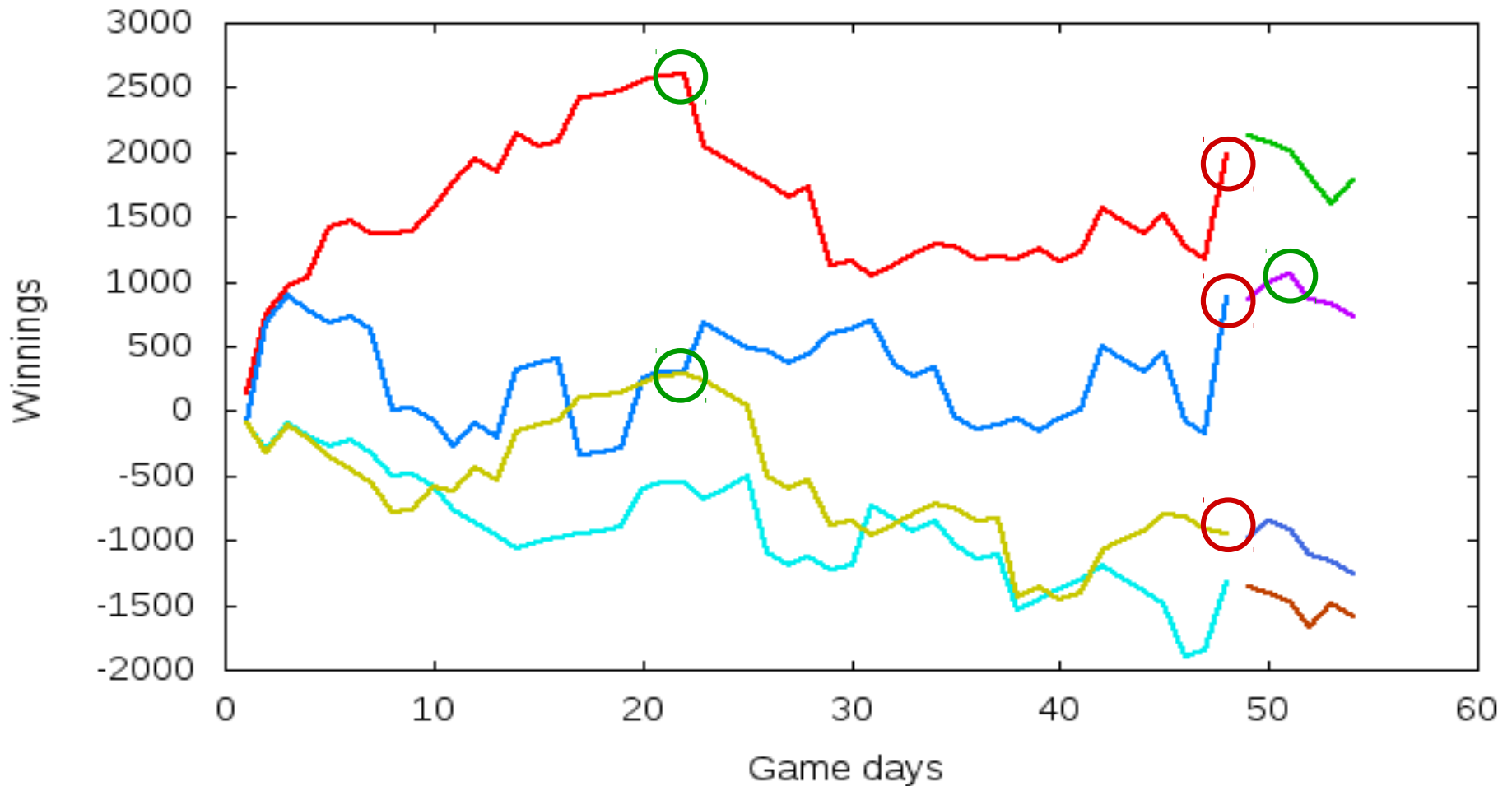
## Distribution of predictions

Classifier	Favs	Dogs	PEs
NB	119	26	14
MLP	103	29	16
RF	111	17	16
SRS	115	18	15



# NFL winning curves

NB (0.6417) ———  
NB post-season (0.6335) ———  
ANN (0.5917) ———  
ANN post-season (0.5896) ———  
RF (0.5792) ———  
RF post-season (0.5737) ———  
SRS (0.5958) ———  
SRS post-season (0.5896) ———



# Can an amateur make money ?

- « Jein »
- ~~March madness : too few matches, too unpredictable~~
- NBA : know when to get in
- NFL : know when to get out

Use Naïve  
Bayes !

# Future work

- Characterize distribution by confidence
- Optimize gains instead of accuracy
- Learn strategies for **which** matches to bet on

Not all  
Favs/Dogs  
are equal !

Exists for  
soccer

**I assume no liability  
for betting losses !**

**Work-in-progress, unpublished things (like NFL  
methodology) :**

**<http://scientificdm.wordpress.com>**

