Learning predictive models for match outcomes in US sports, and using them to bet

Albrecht Zimmermann[0000−0002−8319−7456]

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
albrecht.zimmermann@unicaen.fr

Abstract. Evaluating the accuracies of models for match outcome predictions is nice and well but in the end the real proof is in the money to be made by betting. To evaluate the question whether the models developed by us could be used easily to make money via sports betting, we evaluate three cases: NCAAB post-season, NBA season, and NFL season, and find that it is possible yet not without its pitfalls. In particular, we illustrate that high accuracy does not automatically equal high payout, by looking at the type of match-ups that are predicted correctly by different models. We then put our results to practical use by betting on matches, and find that some observed pitfalls are harder to avoid than others.

1 Introduction

Using advanced sports analytics statistics and Machine Learning (or well-crafted mathematical) models, we can predict match outcomes for a variety of sports – achieving predictive accuracies that are better than chance, a home-field advantage rule-of-thumb, or majority instincts. While this is an interesting (and in our opinion worthwhile) academic exercise, however, the question whether such work is actually useful becomes difficult to avoid. Or, to paraphrase a practitioner of sports betting: "You should compare yourself to betting agencies and see whether you can make money!".

In this work, we intend to do exactly this: using the example of three US sports attracting large betting volumes:

- the main post-season tournament of university (NCAA) basketball – \$ 9.2 billion bet (\$ 262 million legally),
- both regular and post-season of the National Basketball Association (NBA),
- regular and post-season of the National Football League (NFL) – Super Bowl alone \$ 4.1 billion bet (\$ 132 million legally),

we show not only predictive accuracies but also accumulated sports betting outcomes had we used their predictions to consistently place bets in 2015/2016.

We find rather varying outcomes, and, in particular, that very similar accuracies can lead to strongly diverging monetary payoffs. To explore this phenomenon further, we relate this to the way sports betting is handicapped. Finally, we go one step further and employed the derived betting strategy on the NFL and NBA for the 2016/2017 season, as well as the NFL for 2017/2018.

In the following section, we discuss sports betting, and in particular how money-lines should be interpreted and are calibrated. We then discuss our experimental set-up before discussing hypothetical betting outcomes for the 2015/2016 NCAAB post-season, NBA season, and NFL season, respectively, followed by actual betting outcomes for the 2016/2017, and 2017/2018 seasons.

2 Basics of team sports betting

To understand the following discussions, it is necessary to understand money-lines offered by operators of sports betting services (*sports books*), and to have some insight into how those money lines are derived. US sports books offer two ways of betting on match outcomes:

1. *Over-under*, where bettors attempt to correctly foresee the difference between points scored.
2. *Money-line* betting, where bettors attempt to correctly divine the eventual winner of a match.

Given that we have had weak results with trying to predict match scores in the past[1], we ignore the first setting for now, and focus on the second one, which allows us to relate binary predictions to monetary values. Money-lines offered by a sports book for a particular match look as in Table 1.

Match-up	Favorite (FAV)	Underdog (DOG)	FAV-Line	DOG-Line
Detroit Pistons at Atlanta Hawks	ATL	DET	300	240
Utah Jazz at Detroit Pistons	DET	UTH	110	-110

Table 1. NBA money-line examples

For each match, a probable winner (the Favorite) is identified, making the other team the probable loser (Underdog). The associated lines indicate the possible pay-out:

– The **FAV-Line** shows how much money one would **have to bet** to *win* $100,
– the **DOG-Line** how much money one would *win* if one **were to bet** $100.

To make those two settings comparable, we can reformulate the FAV-Line since betting $ 100 would net the bettor $ 10000/FAV-Line. For the first example given in Table 1, Atlanta was considered the favorite, so betting $100 on them and winning would have paid out $33.33. Detroit was expected to lose but if one had bet on them and they had defied predictions, one would have won $240.

Sports books do their best to calibrate those lines, trying to balance two attractions for bettors:

[1] Not yet published.

1. Betting on the favorite is less risky ⇒ has a higher **chance** to pay out.
2. Betting on the underdog and winning will lead to a higher **absolute** pay-out.

Ideally, a match's handicap attracts bettors in such way that the wins that the sports book needs to pay out are offset by the losses of those who bet on the other team (minus some profit for the sports book itself). This can be most clearly seen in the second example in Table 1, a so-called **Pick 'em**. This is a match where the sports book operators do not have enough information to reliably predict one team as winning so betting on either one gives the same pay-off: $ 10000/110 = 90.90. Given a large enough number of bettors, one would expect that roughly half bets on either team and since the sports book pays out $91 for every $100 bet, it would stand to make a profit of 9%.

3 Simulation set-up

Since we are going to use the same general set-up in the succeeding sections, we describe it here.

For each predictive setting, we have collected the money lines for all matches from `http://www.vegasinsider.com/`. The site lists the money-lines offered by the major sports books operating out of Las Vegas, Nevada, which occasionally differ slightly from each other. Additionally, money-lines vary with time, either due to the influx of new information (injuries, player arrests, coaches' announcements), or in reaction to bettors' behavior: too much interest in one team will lead to adjustment in favor of the other one. To avoid undue optimism when evaluating our predictors, we selected the most conservative line for our prediction. If a match is, for instance, listed (1) with FAV-Line=175, DOG-Line=155, and (2) with FAV-Line=165, DOG-Line=145, we will choose

– (1) if our model predicts the favorite to win, since this will pay out **less**, and
– (2) if our model predicts an upset, since we would only get $145 instead of $155.

We use our models' predictions to select on which team to place the bet, and assume that we bet $100 on **every match**[2] in the time period. Correctly predicting a win by the favorite increases the model's winnings by $(10000/FAV-Line)$\$, correctly predicting an underdog's win by DOG-Line$, and correctly predicting the winner of a Pick 'em by 90.90$. Incorrectly predicting a match outcome decreases winnings by $100. For the sake of convenience, we predict matches, and tally up winnings, per day.

The preceding paragraph illustrates an important dynamic – incorrectly predicting is always bad but not all correct predictions are equal:

– Correctly predicting **underdog** wins is the most attractive option and depending on the money-line can balance out several incorrect predictions.
– Correctly predicting **Pick 'ems** still gives a relatively high pay-out.
– Correctly predicting **favorite** wins, on the other hand, needs to happen at a high rate to make up for incorrect predictions.

[2] No matter the odds.

4 2015/2016 NCAAB predictions (and simulated bets)

In our first setting, we consider the NCAAB post-season tournament, also referred to as "March Madness", for the interest and amount of sports betting it generates. This is the smallest of the settings we discuss since the tournament involved only 67 matches.

For US basketball, we base our choices on the evaluation of representations and classifiers reported in [13]. Based on the results reported therein, we use the *Adjusted Efficiencies* pioneered by Ken Pomeroy [9], combined into a weighted average over the season, to encode teams, as well as season-level statistics such as the

- win percentage
- margin of victory – how many more points the team scores in wins than its opponent
- point differential – how many more (or less) points the team scores than its opponent on average

We evaluate three classifiers: *Naïve Bayes* (NB), *Multi-layer Perceptron* (ANN), and a simplified version of Ken Pomeroy's predictor based on the Pythagorean Expectation (KP). We referred to this classifier as "simplified" since we did not estimate the involved coefficients ourselves but based them on the discussions found on his blog. NB and ANN are used in their Weka [6] implementations, with default parameters, except that for NB *Kernel estimator* is set to *true*.

Before discussing the performance of our models, we need to establish the baseline. This means basing ourselves on the money-lines offered by sports books by assuming that we always follow the lead of the money-line. Concretely, if the team designated as favorite wins, we count this as a correct prediction for a **hypothetical "Vegas" model**, if the underdog wins, an incorrect one, with winnings accrued as described above. The main problem for this baseline is posed by Pick 'ems: since the money lines give no indication but we would have to make a prediction, this amounts to flipping a coin for each Pick 'em. In the best case, we, as a bettor, get each of those coin flips right, in the worst case, every single one wrong. Since the difference between getting a Pick 'em right and wrong amounts to \$190.90 per match (the lost gain + the \$100 bet), this leads to a large difference over the course of a season. Since we would assume to get half the coin flips right, we report this as **expected** accuracy/pay-out in Table 2.

w/o Pick 'ems		w/ Pick 'ems (5)					
Accuracy	Pay-out	Best Acc.	Pay-out	Exp. Acc	Pay-out	Worst Acc.	Pay-out
0.7419	30.26	0.7611	484.76	0.7313	7.51	0.6865	-469.73

Table 2. Predictive accuracies/betting pay-outs for "Vegas" for the NCAAB post-season

We can see that always picking favorites would have gotten about 3/4 of the matches right, and paid out approximately $30. Flipping coins on the Pick 'ems can lead to winnings of almost $500 but also to losses of the same magnitude. Especially for so few (five) Pick 'ems, this is a very real risk.

Classifier	NB	ANN	KP
Accuracy	0.6865	0.6417	0.7014
Pay-out	293.52	-605.92	-231.34

Classifier	Favs	Dogs	Pick 'ems (of 5)
NB	39	5	2 (0.4)
ANN	38	2	3 (0.6)
KP	43	0	4 (0.8)

Table 3. Predictive accuracies and betting pay-outs for three predictive models(left), Correct predictions by money-line characterization (right) for the NCAAB.

Results for the predictive models are shown on the left-hand side of Table 3. Two things are immediately noticeable: 1) the relative high predictive accuracies – the KP model performs almost as well as the expected "Vegas" result, and 2) that high accuracy do not translate into a high pay-out. Indeed, while Naïve Bayes performs 1.5 percentage points worse than KP, it shows solid gains (better than using moneylines to bet only on favorites), while KP loses money.

We find some explanation for this phenomenon by looking at the right-hand side of Tab. 3. NB predicts 5 upsets correctly, and even though KP predicts favorite wins and close Pick 'ems better, this makes all the financial difference.

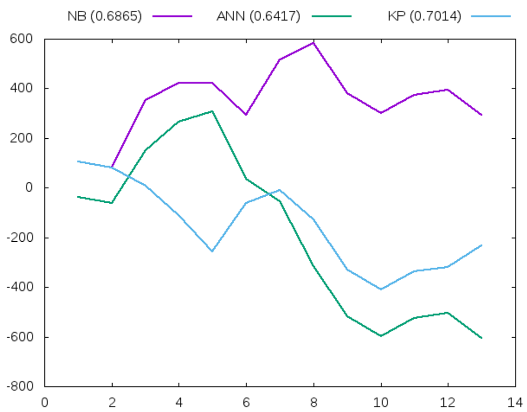


Fig. 1. Classifier winnings over the course of the post-season, NCAAB

The winnings curve for the different classifiers are shown in Figure 1 and shows that the winning behavior is rather erratic. Especially the ANN, which at some point posts winnings similar to the final outcome for NB, drops off into steep loss. But even the NB *could* have returned twice of the final pay-out, a peak that is flanked by losses, however.

5 2015/2016 NBA predictions (and simulated bets)

Our second setting concerns the NBA. We predicted matches for the 2016 regular and post-season. Based on the results in [13], we use NB, ANN, *Random Forest* (RF),[3] as well as the simplified Ken Pomeroy model (KP). Teams were represented by the same statistics as for the NCAAB predictions. We did not predict the first two days of play since at that time a predictor would not have statistics for all teams. As in the preceding section, we establish the baseline, shown in Table 4, both for the regular season, and regular and post-season combined.

Regular season							
w/o Pick 'ems		w/ Pick 'ems (109)					
Accuracy	Pay-out	Best Acc.	Pay-out	Exp. Acc	Pay-out	Worst Acc.	Pay-out
0.7096	-1502,00	0.7356	8407.09	0,6904	-1983,14	0.6461	-12402
Regular + post-season							
w/o Pick 'ems		w/ Pick 'ems (115)					
Accuracy	Pay-out	Best Acc.	Pay-out	Exp. Acc	Pay-out	Worst Acc.	Pay-out
0.7121	-2374.16	0.7375	9125.84	0.6937	-1857.3	0.6492	-12828.81

Table 4. Predictive accuracies and betting pay-outs for "Vegas" for the NBA

Following the money line over the course of the entire season, while ignoring the Pick 'ems, would lead to a very respectable accuracy but also significant monetary loss. At ~1200 matches, the loss per match is only about \$1 yet over the course of the season this accrues. Correctly predicting half the Pick 'ems does of course not improve this by much, even though accuracy would stay high.

Figure 2 plots the development of the different classifiers' winnings over the course of the season, the legend is annotated with predictive accuracies. None of them show a net positive payout, and with the exception of the KP model, they all drop rather low. Notably, they all recover to a certain degree after the mid-point of the season, however, so that one could win money if one could determine *when to start betting.*

The details of the winning curves, showing the difference between the trough and the best result, and its magnitude, can be found in Figure 3. For the ANN and KP, the best result is in the post-season, for NB and the RF in the regular season, even though NB and KP have the same regular season accuracy. Table 5 shows why: while KP strongly outperforms NB when predicting favorites (as for the NCAAB), it underperforms when it comes to Pick 'ems. Pick 'ems are clearly the most difficult matches to predict, and with KP combining three estimated

[3] Which we omitted for the NCAAB, as its accuracy is too weak.

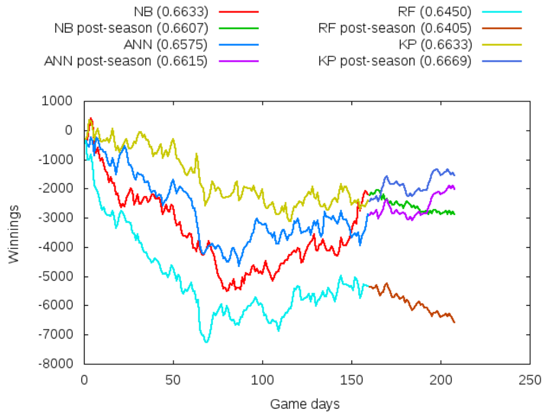


Fig. 2. Classifier winnings over the course of the season, NBA

influences – adjusted efficiencies, the coefficient in the Pythagorean Expectation, and the home-court adjustment – small errors can spiral. The trough-peak difference aligns with the amount of underdog/pick 'em predictions.

Those curves also show something else: If one could start betting at the right time, ~day 75 for NB, or 90 for the ANN, one could in theory win quite a bit of money.

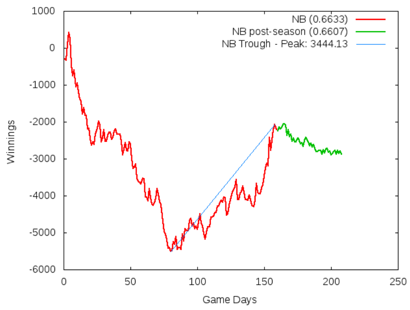
	Regular season			Post-season		
Classifier	Favs	Dogs	Pick 'ems (109)	Favs	Dogs	Pick 'ems (6)
NB	691	57	48 (0.44)	49	5	1 (0.16)
ANN	707	60	22 (0.20)	57	6	0
RF	685	61	28 (0.26)	47	4	0
KP	725	59	12 (0.11)	58	5	0

Table 5. Correct predictions by money-line characterization for the NBA

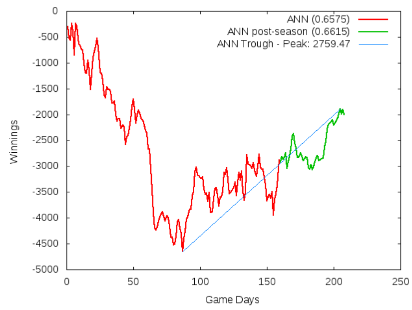
6 2015/2016 NFL predictions (and simulated bets)

The 2015/2016 season marked our first attempt at NFL predictions. As for basketball, the main question to answer concerns team representations. In basketball matches, individual events are *possessions* that lead to either points, or a number of possibly possession-changing events[4]. In American Football, on the other hand, individual events are *Downs*[5] and their outcome is mainly measured in

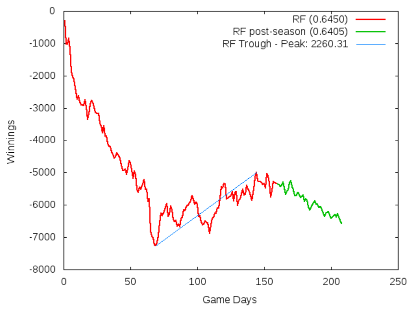
[4] Fouls, turnovers, missed shots
[5] A team has 4 downs to advance 10 yards, after which they get a new set of downs.



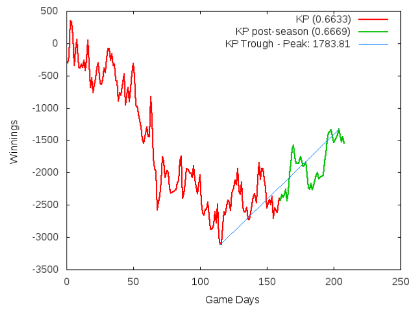
(a) Trough to peak winnings for the NBA season, NB



(b) Trough to peak winnings, ANN



(c) Trough to peak winnings, RF



(d) Trough to peak winnings, KP

Fig. 3. Trough to peak winnings for the NBA season, different predictive models

yards gained (or lost). While the discrete outcomes in basketball can be read off the final box score, the fluctuation of yards in football is less well captured.

To address this, Football Outsiders have proposed *Defense-adjusted Value Over Average* and *Defense-adjusted Yards Above Replacement* [5], both of which consider the outcome of each down in relation to the league-wide average against a particular *defensive alignment*. Since this requires access to and work with play-by-play statistics, we forwent this approach and instead evaluated several other statistics over past seasons:

– Basic Averages – all the statistics available from a typical box score at [1] under "team stats", normalized for 65 possessions, and averaged in a weighted manner (recent games have more weight), both offensively (scored/gained/committed) and defensively (allowed/caused). This follows similar reasoning as possession-based normalizing and averaging in basketball.
– Opp. Averages – same as above but for the opponents that have been played so far. This is intended to help gauge the competition.
– Adjusted Averages – certain offensive and defensive statistics adjusted by mirror statistics of the respective opponents. That is basically the same idea as Ken Pomeroy's adjusted efficiencies [9]. We calculated those statistics in the same manner as in our work on basketball [13].

- SRS – the "simple rating system" information (SRS, SoS) as described at [10], with the difference that the averaging is weighted, i.e. not divided by number of matches.

The first surprise was that the feature selection techniques contained in the Weka package did not seem to work at all: no consistent set of features could be found across seasons (apart from *rushing yards allowed*), and the set of features was often very small. To address this, we chose to simply evaluate all presentations for several predictors – ANN, NB, RF – for seasons 2009-2014 (Table 6).

	Encoding	2009	2010	2011	2012	2013	2014
ANN	Adjusted Avg	0.5863	**0.5663**	0.5863	0.5783	0.5422	0.5703
	Adjusted+SRS	0.5542	0.5462	0.6024	0.5743	0.5663	**0.6546**
	Basic Avg+Opp. Avg	**0.6104**	0.5341	**0.6145**	0.5823	0.5261	0.5984
	Basic+SRS+Opp.	0.5904	0.5261	0.5823	**0.6225**	0.5622	0.6386
	Basic+SRS	0.5663	0.5181	0.5622	0.5663	**0.5904**	0.5904
	Basic Avg	0.5542	0.5020	0.5502	**0.6225**	0.5703	0.5783
	All of the above	0.5863	0.5462	0.6024	0.5904	0.5743	0.6345
NB	Adjusted Avg	0.6305	0.6024	0.6145	0.6265	0.5582	0.3815
	Adjusted+SRS	0.6426	0.6024	0.6305	0.6345	0.5582	0.4016
	Basic Avg+Opp. Avg	0.5863	0.5542	0.6667	0.6064	0.5984	0.6104
	Basic+SRS+Opp.	0.6024	0.5422	0.6627	0.6064	0.5944	**0.6305**
	Basic+SRS	0.6466	**0.6064**	0.6787	0.6024	0.6185	**0.6305**
	Basic Avg	**0.6586**	0.6024	**0.6867**	0.6104	**0.6305**	0.6265
	All of the above	0.5944	0.5622	0.6787	**0.6386**	0.5703	0.4779
RF	Adjusted Avg	0.6064	0.5181	0.6185	0.5622	0.5703	0.6104
	Adjusted+SRS	0.6024	0.5261	0.6345	**0.6064**	0.5542	0.5823
	Basic Avg+Opp. Avg	**0.6305**	0.5422	0.6024	0.5863	0.5582	0.6305
	Basic+SRS+Opp.	0.5502	0.5261	0.6064	0.5462	**0.5904**	0.5984
	Basic+SRS	0.6064	**0.5823**	0.5542	0.5904	0.5622	0.6185
	Basic Avg	0.6145	0.5261	0.6064	0.6024	0.5663	**0.6546**
	All of the above	0.6265	0.5382	**0.6466**	0.5301	0.5301	0.6064

Table 6. Evaluation of several team representations for NFL prediction

There is no clear trend but it did often seem like a good idea to involve the basic averages. We initially settled on Basic Averages for NB, Basic+Opp. Averages for ANN, and since RF did not present any trend at all, we tried RF with Adjusted Averages. After additional evaluation during the 2015/2016 season, we found better performances by using Basic+Opp. for NB, and Adjusted for ANN and RF.

We also evaluated the SRS. We did not predict the first week's matches since for those matches, since we do not have statistics for the teams at that time. Again, we need to establish the baseline (see Table 7).

The baseline again shows consistent behavior: the accuracy is relatively high but if one follows money-line predictions one loses – not much per individual game but quite a bit in the aggregate.

Regular season							
w/o Pick 'ems (28)		w/ Pick 'ems					
Accuracy	Pay-out	Best Acc.	Pay-out	Exp. Acc	Pay-out	Worst Acc.	Pay-out
0.6367	-1443.45	0.6791	1102.49	0.6208	-1570.11	0.6375	-4242.71
Regular + post-season							
w/o Pick 'ems		w/ Pick 'ems (29)					
Accuracy	Pay-out	Best Acc.	Pay-out	Exp. Acc	Pay-out	Worst Acc.	Pay-out
0.6441	-1215.69	0.6852	1420.68	0.6294	-1251.92	0.5697	-4115.42

Table 7. Predictive accuracies and betting pay-outs for "Vegas" for the NFL (2015/2016 season)

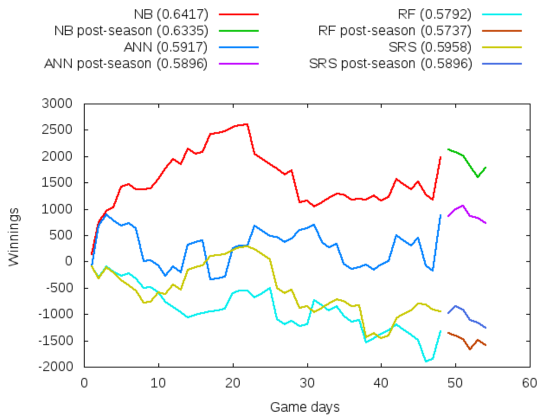


Fig. 4. Classifier winnings over the course of the season, NFL

The results of the predictors, shown in Figure 4, are very interesting. The first thing to notice is that NB, using rather straight-forward statistics, achieves comparative accuracy to "Vegas" and a much better pay-out. In fact, its pay-out is better than that for the best-case "Vegas"-scenario in the regular season. Even the ANN, with much lower accuracy, achieves a good pay-out. Additionally, we again see the influence of which picks to predict correctly at play: even though ANN and SRS have very similar accuracies, betting according to SRS would be a clear loss, and the difference can be explained by the fact that the ANN trades off accuracy on favorites against accuracy on underdogs (Table 8).

A final monetary realization is that each predictor reaches a high point that comes before the end of the season. In fact, following NB all to the end of the regular season would mean forfeiting more than \$ 600, with losses for all models in the post-season. While for NBA predictions it seems to be important to know when to get *in*, for the NFL is important to know when to get *out* – a decision that might be slightly easier to make.

	Regular season			Post-season		
Classifier	Favs	Dogs	Pick ’ems (28)	Favs	Dogs	Pick ’ems (1)
NB	115	25	14 (0.5)	4	1	0
ANN	98	29	15 (0.54)	5	0	1 (1.0)
RF	107	16	16 (0.57)	4	1	0
SRS	111	18	14 (0.5)	4	0	1 (1.0)

Table 8. Correct predictions by money-line characterization for the NFL

7 Use cases – actual betting

7.1 2016/2017 seasons

Based on our simulated results, we employed the following betting strategy:

- For the NFL, we began betting with the first match of the second week. Its peak was reached on 05/11/2015, so we would stop after the first week in November 2016, or if we were at risk of losing more than €1500 (€100 · 15 matches per week).
- For the NBA, we began betting with the last day of matches before the All-Star break (16/02/2017), with the intention to continue until the end of the regular season.

Given the results of our simulation, and particular its performance regarding underdog and Pick ’em-wins, we used Naïve Bayes predictions. We placed our bets on the website `1xbet.com`. Because there were restrictions in place regarding the transfer of funds online, we only bet €50 on each match on the second Sunday of the NFL season (18/09/2016) – our second day of betting – and €100 on all matches before and after.

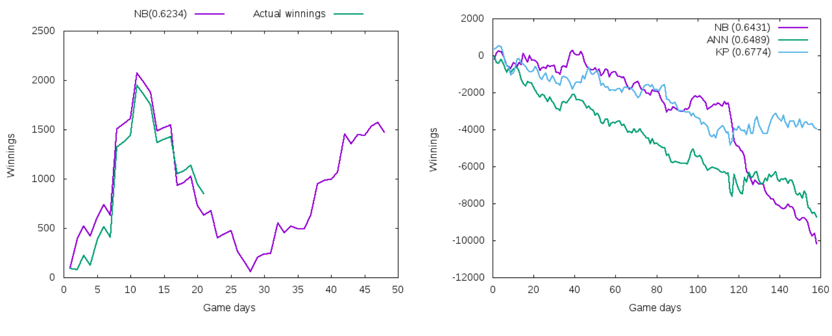


Fig. 5. Betting outcomes for the 2016/2017 season, NFL (left), and NBA (right)

The left-hand side of Figure 5 shows the outcomes for 2016/2017, with the curve labeled "NB" indicating the winnings according to Vegas moneylines[6], and the curve labeled "Actual Winnings" the payout from `1xbet.com`. The peak that we had observed in 2015/2016 appears again but unfortunately **much earlier** (09/10/2016). Had we stopped at that moment, we would have won about €2000. Instead, we continued until 31/10/2016, eventually winning ≈€850, from an initial investment of €1200.[7] A notable development is that the actual winnings at some point increased over those according to the Vegas moneylines. It is our understanding that online sports books eventually individualize odds to discourage too-successful bettors from certain bets but since we bet indiscriminately, this had no effect on our behavior. At the peak, we had bet on sixty matches, and NB had an accuracy of 68.33%, and had correctly predicted **seventeen** underdog wins! When we stopped, the accuracy had dropped to 59.78%, with only two more correctly predicted underdog wins. However, as the figure shows, there was a second (smaller) peak towards the end of the season. Over the course of the entire season, NB with averages + opponent averages achieved an accuracy of 0.6234, and had predicted twenty-seven underdog wins, or almost a quarter of all correct predictions.

This partial success is however tempered by our experience with NBA betting. The right-hand side of Figure 5 tracks the performance of our betting strategy (NB) as quantified by Vegas money lines. Not only did we not experience a climb out of the trough but the classifier quickly lost money. We stopped the experiment after only nine days of betting, after having lost the entire winnings of the NFL experiment. Apart from the fact that all classifiers continuously lost money, the NB fell off a cliff after 3/4 of the season. Notably, the other classifiers would not have really improved the outcome.

7.2 2017/2018 NFL season

The 2016 season's NBA results discouraged us but encouraged by the results of the 2015/2016 simulation, and 2016/2017 betting results, we also bet during the 2017/2018 season on the NFL.

As Figure 6 shows, this has not been rewarded. All classifiers (including the NB(Avg+OAvg) which we used for betting) lost heavily early on. This has a very concrete effect on the bets: losing 600 € in the first week means that one cannot continue with the 100 € per match strategy, unless one adds significant additional funds. Instead, we either split remaining money among all matches or focused on high-yield bets. As a result, our actual winnings fall back behind the simulated results. As the plot also shows, it is not clear which classifier and which representation to choose.

[6] We show this for comparison's sakes since `1xbet.com`'s moneylines were slightly different.

[7] We withdrew ≈ €2050 from the website.

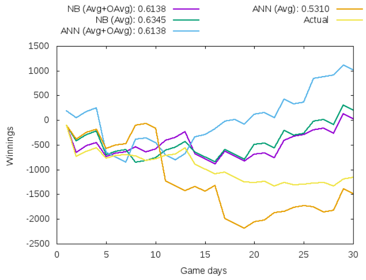


Fig. 6. Winnings for the 2017/2018 NFL season

8 Related work

The majority of work on sports betting is concerned with the question of how to exploit the difference between implied win probabilities and those predicted by different models, a question that we do not touch upon. They also notably usually *do not* actually wager any money using the proposed strategies.

Constantinou [2] exploits a Bayesian network model in simulating betting on the English Premier League. They exploit the difference between predicted probabilities and implied probabilities according to betting odds, and report both of their simulated betting strategies showing significant betting gains.

Langseth [7] also exploit the difference between predicted and implied probabilities in simulating betting on soccer. Evaluating a number of different betting strategies, and remarkably enough, they report that almost all combinations of predictive models and betting strategies would win money (in excess of 10% of the wagered amount). Looking at details, however, they find that a large part of the winnings are due to two underdog wins, and that stopping early would have improved pay-outs.

Pfitzner *et al.* [8] consider NFL football, using regression to predict over-under results. They formulate a kind of adjusted performance statistics, and evaluate different betting strategies. They state that betting strategies must have a win rate of at least 52.4% to beat the built-in sport book profit margin. Notably, betting the same amount on all matches leads to a win rate of 55%, and is therefore viable, in their simulation.

Pessemier *et al.* [3], contrary to our approach, exploits the confidence of model predictions to evaluate a number of different betting strategies, one of which consists of alway betting on underdogs at home. The simulate the results of their approach on a number of different European soccer leagues, and find that different leagues might require different strategies. They find that betting the same amount on every match has the highest possible pay-off.

9 Conclusion and outlook

The answer we intended to explore in this paper was whether it is possible to use match outcome predictions for US sports for betting. Given the evidence presented, the answer has to be "Maybe". We explored three settings, the NCAAB playoffs, NBA, and NFL regular and postseasons. In each case, we built a model using a set of statistics that is relatively easy to acquire (downloadable from publicly accessible websites), and to process. Once a model has been established, it can be used to place bets in a straight-forward manner by betting an equal amount on each match.

We used our models to simulate betting for the 2015/2016 season, and found that having an over-all positive return is possible using the Naïve Bayes classifier. However, the NCAAB post-season contains few matches, leading to rather volatile pay-out. In the NBA, one can win but only after figuring out when to start betting. In the NFL results, finally, straight-forward use could indeed lead to a decent pay-off (admittedly, not attractive to professional gamblers), especially if one stops early enough.

We have also illustrated one of the aspects that make a predictive model more or less well-suited for sports betting, by considering what kind of matches models predict well. In particular, a model that is not very strong in correctly predicting favorites but performs well for Pick 'ems, or even better, matches won by underdogs, would be a particular attractive tool, even if its straight-up accuracy is not impressive.

Based on the conclusions from our simulations, we put our work into practice, betting on the 2016/2017 NFL season (stopping early), and NBA season (beginning at roughly the mid-point of the season). We found that the peak in NFL betting does not directly transfer from season to season but that the general behavior – an early peak with a good payoff, followed by a decline in winnings – could in fact be exploited. For the NBA, our simulation results did not transfer, leading to a quick loss of the NFL gains. The 2017/2018 NFL season, finally, showed that significant losses are possible. The main reason for this is chance: a surprising amount of professional US sports matches are decided by chance. Past work has tried to characterize that phenomenon based on particular distributions or upsets by underdogs [4]. Yet underdogs are in the eye of the beholder, and a recent work instead clustered teams and evaluated relative win probabilities [14]. They found that **half** of all match outcomes in college basketball are due to chance and given the relatively low scores in the NFL, it is to be expected that chance plays an even larger role there.

We intend to explore the question of why models perform well or not further by relating models' performance to evaluations based on lift-charts and ROC-like discussions. The final goal would of course be to shift the training of predictive models: away from maximizing predictive accuracy and towards maximizing payouts, which means predicting border-line cases correctly instead of easy ones. A different direction consists of proposing which matches (not) to bet on and/or how much to bet, as has been done in [7, 11] for soccer, or in [12] for American Football. Possible approaches include leveraging game theoretic approaches or reinforcement learning.[8] It should be noted, however, that such strategies would make our betting choices susceptible to the adjusted odds mentioned in Section 7.

References

1. Pro-football reference. http://www.pro-football-reference.com
2. Constantinou, A.C.: Bayesian networks for prediction, risk assessment and decision making in an inefficient association football gambling market. Ph.D. thesis, Queen Mary, University of London (2013)
3. De Pessemier, T., De Deyn, B., Vanhecke, K., Martens, L.: Recommendations for sports games to bet on. In: Second Workshop on Recommendation in Complex Scenarios (ComplexRec 2018), in conjunction with the 12th ACM Conference on Recommender Systems (RecSys 2018). pp. 8–12 (2018)
4. E. Ben-Naim, F. Vazquez, S.R.: Parity and predictability of competitions. Journal of Quantitative Analysis in Sports **2**(4) (2006)
5. Football Outsiders: Methods to our madness. http://www.footballoutsiders.com/info/methods
6. Frank, E., Witten, I.H.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann (1999)
7. Langseth, H.: Beating the bookie: A look at statistical models for prediction of football matches. In: Jaeger, M., Nielsen, T.D., Viappiani, P. (eds.) Twelfth Scandinavian Conference on Artificial Intelligence, SCAI 2013, Aalborg, Denmark, November 20-22, 2013. Frontiers in Artificial Intelligence and Applications, vol. 257, pp. 165–174. IOS Press (2013). https://doi.org/10.3233/978-1-61499-330-8-165, http://dx.doi.org/10.3233/978-1-61499-330-8-165
8. Pfitzner, C., Lang, S., Rishel, T.: Factors affecting scoring in nfl games and beating the over/under line. The Sport Journal (2014)
9. Pomeroy, K.: Advanced analysis of college basketball. http://kenpom.com
10. Pro-Football-Reference.com: A very simply rating system. http://www.pro-football-reference.com/blog/?p=37
11. Snyder, J.A.L.: What Actually Wins Soccer Matches: Prediction of the 2011-2012 Premier League for Fun and Profit. Master's thesis (2013)
12. Warner, J.: Predicting margin of victory in NFL games: Machine learning vs. the Las Vegas line. Tech. rep. (2010)
13. Zimmermann, A.: Basketball predictions in the NCAAB and NBA: similarities and differences. Statistical Analysis and Data Mining **9**(5), 350–364 (2016)
14. Zimmermann, A.: Chance and the predictive limit in basketball (both college and professional). In: International Conference on Discovery Science. pp. 599–613. Springer (2023)

[8] We thank prior reviewers for this suggestion.