Machine Learning the Harness Track: Crowdsourcing and Varying Race History

Robert P. Schumaker

Management Information Systems
Central Connecticut State University, New Britain, Connecticut 06050, USA
rob.schumaker@gmail.com

Word Count: 7,849

Abstract

Racing prediction schemes have been with mankind a long time. From following crowd wisdom and betting on favorites to mathematical methods like the Dr. Z System, we introduce a different class of prediction system, the S&C Racing System that derives from machine learning. We demonstrate the S&C Racing system using Support Vector Regression (SVR) to predict finishes and analyzed it on fifteen months of harness racing data from Northfield Park, Ohio. We found that within the domain of harness racing, our system outperforms crowds and Dr. Z bettors in returns per dollar wagered on seven of the most frequently used wagers: Win $1.08 return, Place $2.30, Show $2.55, Exacta $19.24, Quiniela $18.93, Trifecta $3.56 and Trifecta Box $21.05. Furthermore, we also analyzed a range of race histories and found that a four race history maximized system accuracy and payout. The implications of this work suggest an informational inequality exists within the harness racing market that was exploited by S&C Racing. While interesting, the implications of machine learning in this domain shows promise.

Keywords: Business Intelligence, Data Mining, Support Vector Regression, Harness Racing, S&C Racing System, Crowdsourcing, Dr. Z System

1. Introduction

Prediction, the art of divining future events has always had a certain appeal. From the earliest of times, man cast lots to determine outcomes. Within these early times, lots[1] of sticks and pebbles were given interpretations where an even number indicating a positive outcome and an odd number was viewed as negative. Gradually these divinations grew more complex with rituals, sacrifice and complicated interpretations [1].

Today, the grip of prediction still has its appeal with gamblers and academics alike. While no longer considered an art but more of a science, prediction and probability are better understood, but are still complex by design. The most difficult aspect of prediction rests within identifying the most relevant parameters. Critical parameters are sometimes difficult to identify/measure, are constantly changing or may not have been fully explored. The inability to correctly identify the most relevant parameters can sometimes lead to crippled systems relying on unimportant data, or worse, not based on sound science (e.g., basing predictions on the color of a horse).

Racing, like so many other domains including the stock market, trades on information. In harness racing it is assumed that all relevant information is public. Therefore everyone should have a fair chance of achieving success. Instead, what has been found is that information markets all have inequalities within them, either through withheld information or a human tendency to discount certain information or weight it incorrectly. These informational inequalities lead to arbitrage opportunities. The key to unlocking successful prediction rests on making unbiased decisions using measurable parameters of the most critical values.

To a large extent, individual bettors will base wagering decisions on a mixture of quantitative and qualitative methods as well as instinct. Before making a wager, a bettor will

---

[1] An Old English term for an object used in random selection.

typically read all the information on the race card and gather as much information about the race participants as possible. They will also examine data concerning a horse's physical condition, how they have performed historically, breeding and bloodlines, trainer or owner, as well as odds of winning.

Automating this decision process and removing the biases using machine learning may yield as equally good results as greyhound racing, which is considered to be the most consistent and predictable form of racing [2]. Consistency lends itself well to machine learning algorithms that can learn patterns from historical data and apply itself to previously unseen racing instances. The mined data patterns then become a type of arbitrage opportunity where an informational inequality exists within the market. However, like other market arbitrages, the more it is exploited the less the expected returns, until the informational inequality returns the market to parity.

Our research motivation is to demonstrate a decision support system that can learn from historical harness race data and leverage the information arbitrage through its predictions. Since nothing of this nature has been performed before in harness racing, we ground ourselves in using similar techniques used in greyhound racing research and then compare our artifact against bettors and other successful wagering strategies; testing how well such a system performs.

The rest of this paper is as follows. Section 2 provides an overview of literature concerning prediction techniques, algorithms and common study drawbacks. Section 3 presents our research questions. Section 4 introduces the S&C Racing system and explains the various components. Section 5 sets up the Experimental Design. Section 6 is the Experimental Findings and a discussion of implications. Finally Section 7 delivers the conclusions and limitations of this stream of research.

2. Literature Review

Harness racing is a fast-paced sport where horses pull two-wheeled sulkies (e.g., carts) with jockeys inside. Quite popular in North America, there are 42 sanctioned USTA tracks mostly throughout the Midwest. Races are completed with two different gaits; pacing or trotting which has to do with leg positioning. In pacing, leg movements are laterally coordinated, the left side moves together and opposite of right side movement. In trotting, leg movement is coordinated diagonally. A majority of harness races are pacing races because of the faster speeds.

2.1 Predicting Outcomes

Participants can use a wide variety of prediction techniques to divine potential race outcomes. Many of the techniques can be reduced to three major areas; market efficiency, mathematics or data mining [3]. Market efficiency focuses on the psychological aspects of a bettor and how motivations can lead to informational inequalities. While typically applied to stock markets where information inequalities can arise from public versus private information, the same concept can be applied to other information-centric markets. This area relies on using statistical tests, models of human behavior, historical forecasting and crowdsourcing.

A second branch of prediction techniques is that of mathematical models where problem sets of variables are reduced to a solvable model. Examples include using joint probabilities, the Dr. Z System and the mathematics of streaky behavior.

A third major branch of prediction techniques is data mining. In this branch, algorithms are harnessed to learn from historical data and generalize to new instances. Examples include performing simulations, using artificial intelligence and machine learning algorithms [3].

2.1.1 Market Efficiency

Market Efficiency is all about the movement and use of information within a tradeable environment. Market Efficiency includes the use of statistical tests, conducting behavioral modeling of trading activity and forecasting models of outcomes to create rules from observation [4]. Crowdsourcing, or using the wisdom of crowds, is another form of market efficiency where groups of individuals perform forecasts on provided information and results are averaged for use as a predictive tool [5]. Since Market Efficiency is such a broad field, we constrain our discussion to the narrowly scoped domain of sports gaming.

In statistical testing, it is assumed that outcomes will mirror expectations and no betting strategy should win in excess of 52.4% provided that all participants share the same information [4]. Deviations from these expectations could indicate the introduction of new or privately held information.

In Behavioral models, bettor biases are tested to determine predictable decision-making tendencies. Perhaps the best known behavioral model is the longshot bias where a bettor will over-value horses with higher odds to offset losses [6]. Gamblers tend to favor the low probability, high payout combinations for luck, entertainment or desperation, but it has never been found to yield sustainable positive returns on combination bets [7]. It has been argued that betting on favorites should be as profitable as betting on longshots [8], however in practice this is not the case. The longshot bias is not constrained to one particular sport and has been observed in boxing, cricket, horse racing, snooker and tennis [9].

Another Market Efficiency tool is forecasting where models of historical data (e.g., seasonal averages, basic statistics and prior outcomes) are extrapolated and tested against current data [4]. This technique was found to be too simplistic and a poor predictor of future activity.

Crowdsourcing is another tool of Market Efficiency where an average of crowd forecasts is used to predict future events [5]. In harness racing, this behavior translates into betting on the favorites. This technique has been found to be fairly accurate and a reliable indicator of expectations. In a study of UFC fights, crowds were able to better predict wins (85.7%) than bookies (67.6%) [10]. In a study of German Premier League soccer, crowds were found to be more accurate than bookies [11]. In a study of FIFA World Cup 2006, crowds were better able to predict winners than rankings or random chance [12]. Although crowdsourcing has been shown to be an effective prediction tool, critics maintain that it could be susceptible to the free-riders problem where certain bettors may simply follow the crowd favorite themselves rather than evaluate the data independently [13]. This could lead to an over-emphasis on the crowd favorite, more so than what the data should demand.

2.1.2 Mathematics

Mathematics in sports racing focuses on problem sets of observable variables and reducing them to a solvable model. It differs from Market Efficiency where the focus is on market information (public, semi-public and private information) amongst participants. This branch of prediction techniques include the Harville formulas, the Dr. Z System and streaky player performance [14].

Harville formulas [15] are a collection of equations that establish a rank order of finish by using combinations of joint probabilities [16]. It is believed that in certain instances the odds can be over-estimated on favorites leading to an arbitrage opportunity.

A derivative work of the Harville formulas is the Dr. Z System. In this system, a potential gambler is leveraging odds over-estimation by waiting 2 minutes before the race and making Place wagers (i.e., the horse will finish in 2nd place or better) on those with a win

frequency to place frequency greater than or equal to 1.15 and betting Show (i.e., the horse will finish in 3rd place or better) on races with a win to show frequency greater than or equal to 1.15 [17]. This system proved successful during the 1980s and received considerable attention from academics and gamblers alike. Follow-up studies later found that bettors were effectively arbitraging the tracks (over-using the system) and any opportunity for gain using Dr. Z was mitigated [18].

In streaky behavior, player performance is analyzed for the so-called "hot-hand" effect to determine if recent player performance influences current performance [19]. The belief is that a player performing statistically better than average, will continue to do so in the near future. While Tversky and Gilovich studied the phenomenon in basketball and did not find evidence of streaky behavior [19], academics studying baseball found just the opposite. In a study modeling baseball player performance, it was found that certain players exhibit significant streakiness, much more so than what probability could allow [20].

2.1.3 Data Mining

While mathematics and statistics are the underlying foundation of data mining, they are very different from one another. Statistics are typically used to identify an interesting pattern in a signal and allow for theory testing of hypotheses on known relations [21]. Without knowing what to look for, researchers using statistics alone cannot explain an observed relationship; that is the purpose of data mining [22]. As an example, a particular baseball batter may have a substantial batting average. Statistics can tell us the value but not answer why that is the case. If we knew what we were looking for, or embarking upon an exploratory analysis of the data, we may arrive at the elevated batting average on a 3-2 pitch count with 2 outs and runners in scoring position while facing a left-handed pitcher. Data mining can identify this important nugget from

the data, but statistics would require manual iteration and an idea of what to look for in order to arrive at the same conclusion. Therefore it could be said that data mining provides more explanatory power than statistics. Data Mining can be broken into three areas; Simulations, Artificial Intelligence and Machine Learning.

Statistical simulations involve the imitation of new game data by using historical data as a reference. Once constructed, the simulated play can be compared against actual game play to determine the predictive accuracy of the simulation. Entire seasons can be constructed to test player substitution or the effect of player injury.

Artificial Intelligence differs from other methods by applying a heuristic algorithm to the data. This approach attempts to balance out statistics by implementing codified educated guesses to the problem in the form of appropriate rules or cases. Heuristic solutions may not be perfect, however, the solutions generated are considered adequate [22].

The third area, machine learning, uses algorithms to learn and induce knowledge from the data [23, 24]. Examples of these algorithms include both supervised and unsupervised learning techniques such as genetic algorithms, neural networks and Bayesian methods. These techniques can iteratively identify previously unknown patterns within the data and add to our understanding of the composition of the dataset [22]. These systems are better able to generalize the data into recognizable patterns [25].

Neural networks are widespread within sport prediction studies. With neural networks, dataset patterns are learned and hidden trends can be exploited for a competitive advantage. Other machine learning techniques include genetic algorithm, the ID3 decision tree and the regression-based variant of the Support Vector Machine (SVM) classifier, called Support Vector Regression (SVR) [26]. SVM is a classification algorithm that seeks to maximally separate high

dimension data while minimizing fitting error. This technique was used in a similar context to predict stock prices from financial news articles [27] and credit ratings [28]. Within these studies, it was found that the SVR algorithm was able to exploit arbitrage opportunities that were the result of market inefficiencies with respect to the information present.

2.2 Racing Prediction Studies

Predictive algorithms have been adopted successfully in non-traditional sports, such as greyhound and thoroughbred racing. These types of predictions generally involve machine learning techniques to train the system on the various data components, feed in new data and then extract predictions from it. The highlights of several studies are presented below.

2.2.1 Neural Networks and ID3

In a prior study of greyhound races, Chen et. al. tested an ID3 and Back Propagation Neural Network (BPNN) on 100 races at Tucson Greyhound Park [2]. They further limited themselves to ten race-related variables over a seven race history. These ten race-related variables on a seven race history were given to Chen et. al. by greyhound domain experts and were never systematically explored to ensure a maximal return.

In Chen et. al.'s work they made binary decisions as to whether or not each greyhound would win based on historic race data (Win versus not Win). If a dog was predicted to finish first (Win), the system would make a $2 wager. Their ID3 decision tree was accurate 34% of the time with a $69.20 payout while the BPNN was 20% accurate with a $124.80 payout. This disparity in decreased accuracy and increased payout is justified by arguing that the BPNN was selecting longshot winners. By doing so, accuracy would decrease but higher payouts could be gained because of the longer odds. When comparing machine learning techniques to track experts, the experts managed a dismal 18% accuracy with a payout loss of $67.60. It was

9

speculated that the system was taking advantage of informational inequalities by successfully selecting longshot wagers more often than chance, however, given the black-box nature of BPNN, it is hard to be certain.

2.2.2 Neural Networks and Betting Engines

In a follow-up study that took Chen et. al.'s variables and expanded them to 18, Johansson and Sonstrod also used a BPNN but also investigated the effect of more exotic wagers such as Quiniela (i.e., selecting the first two dogs to finish in any order) and Exacta (i.e., selecting the first two dogs to finish in order) [29]. Their study on 100 races at Gulf Greyhound Park in Texas found 24.9% accuracy for Wins and a $6.60 payout loss. This seemingly improved accuracy came at the cost of decreased payout as compared to Chen et. al. and would imply that either the additional variables or too few training cases (449 as compared to Chen's 1,600) hampered the ability to identify longshots. However, exotic wagers did better. Quinielas had 8.8% accuracy with a $20.30 payout, while Exacta had 6.1% accuracy with a $114.10 payout.

2.2.3 Discrete Prediction

In another study that focused on using discrete numeric prediction rather than binary assignment, Schumaker and Johnson used Support Vector Regression (SVR) on Chen et. al.'s 10 performance-related variables [30]. Their study of 1,953 greyhound races across the US managed a 45.35% Win accuracy with a $75.20 payout. To maximize payout, AZGreyhound had 23.00% Win accuracy with a $1,248.40 payout. They found the same tradeoff between accuracy and payout.

2.2.4 Neural Networks on Thoroughbreds

In a study of thoroughbred racing, Williams and Li measure 8 race performance variables on 143 races and built a BPNN for each horse that raced [31]. This differed from other BPNN studies that created one network for all races. The system did manage 74% accuracy in selecting a winner.

2.3 Research Gaps

From our investigation, we noticed several gaps in the literature. The first of which is a lack of machine learner study within the domain of harness racing. Although there are a scant few studies within the related race areas of greyhound and thoroughbred racing, harness racing may be similar enough that techniques used in similar domains can be ported to harness racing.

Another gap was a lack of study of machine learners versus the wisdom of crowds. Several studies offered insight between crowds and experts, but none could be found that explored how well a machine learning platform could perform versus crowd wisdom. Perhaps a machine learning algorithm can find patterns that others miss.

A third gap was that nearly every racing study discussed has relied on the ten race-related variables and a seven race history which was derived from an interview of human domain experts in Chen et. al. (1994). The use of these ten variables and seven race history has never been examined to our knowledge to determine if they are ideal. We take this opportunity to investigate what amount of race history can lead to optimal results and why.

3. Research Questions

From our analysis we propose the following research questions.

♦ *Can a Machine Learner predict Harness races better than established prediction methods?*

Crowdsourcing and Dr. Z methods have been well established within the racing domain. However, both of these methods are susceptible to human biases and risk avoidance tendencies. Machine learning is devoid of these human characteristics and should be able to outperform the established prediction methods in a bias-free decision-making environment.

♦ *How important is race history to a machine learner?*

Prior research has almost exclusively relied upon a seven race history to make predictions. We question whether this amount of race history is optimal within the harness racing domain. Perhaps a differing amount of race history will provide better performance.

♦ *What wager combinations work best and why?*

Most prior studies were fixated on Win-only wagers. While Win is still an important aspect in sports wagering, perhaps wagers on Place and Show may be more profitable. Likewise, more exotic wagering types might hold some promise as well. By looking at maximizing the return per wager, we can find the answer.

4. System Design

To address these research questions, we built the S&C Racing system shown in Figure 1.
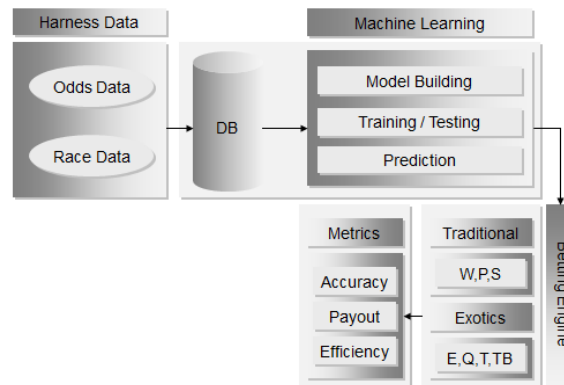

Figure 1. The S&C Racing System

The S&C Racing system consists of several major components: a web scraper/parser to gather the online odds and race history from race programs, the machine learning tool that takes different models and learns the patterns, a betting engine to make different wagers and evaluation metrics to measure system performance. For odds data, harness track odds are pari-mutuel where the track sets the odds to balance the amount of money transfer from losing to winning wagers, minus a commission. Thus if a particular horse is the favorite and is heavily bet upon, the odds are decreased, which decreases payout. To offset favorite betting, the track will increase odds on less favored horses to give bettors an incentive to wager longshots.

These odds are made for a variety of wagers aside from just Win where the bettor receives a payout only if the selected horse comes in first place. Place produces two differing payouts depending upon whether the selected horse comes in either first or second place. A Show bet has three differing payouts that depend on whether the selected horse comes in first, second or third place. These differing payouts are dependent upon the odds of each finish. For exotic wagers, an Exacta bet receives a payout by successfully picking both the first and second place horse. A Quiniela wager is like an Exacta except the order of finish does not matter, only that the selected two horses finish within the top two spots. Trifecta, or Trifecta Straight, is where the bettor wagers on the first three horses, in order. For a Trifecta Box wager, the bettor still wagers on the first three horses, but the order does not matter as long as all three finish within the top 3 spots.

The other system input is the race program which contains historical performance data on each horse. There are generally 14 races per program where each race averages 8 or 9 entries. Each horse has specific data such as name, driver, trainer, color, sire and dam. Race-specific information includes gait (pacing versus trotting), race date, track, fastest time, break position,

quarter-mile position, stretch position, finish position, lengths won or lost by, average run time and track condition.

From this data we create specific models to train the S&C Racing system. The system is tested with different wagers and results are tested along three dimensions of evaluation: accuracy, payout and efficiency. Accuracy is the number of winning bets divided by the number of bets made. Payout is the monetary gain or loss derived from the sum of wagers. Efficiency is the payout divided by the number of bets.

5. Experimental Design

For our collection we automatically gathered data of prior race results and wager payouts at Northfield Park; a USTA sanctioned harness track outside of Cleveland, Ohio. We further chose a study period of October 1, 2009 to December 31, 2010 and divided it into two periods; training (October 1, 2009 to September 30, 2010) and testing (October 1, 2010 to December 31, 2010). Fifteen months were selected because it gave us a comparable amount of training races/cases as prior studies. In all, we gathered 2,558 useable training cases covering 309 races and tested our system on 91 testing races covering 770 testing cases. By comparison, Chen et. al. (1994) used 1600 training cases from Tucson Greyhound Park, Johansson and Sonstrod (2003) used 449 training cases from Gulf Greyhound Park in Texas and Schumaker and Johnson used 41,473 training cases from across the US.

Part of the challenge in constructing this system was in maintaining consistency with Chen et. al.'s approach which required a race history of the prior seven races. Each usable race needed a complete 7 race history of each horse participating, otherwise the entire race was discarded. Since new entries would arrive in the Northfield market frequently, only a portion of all races during this period could meet our stringent requirement.

Following the work of Chen et. al. (1994), we limited ourselves to the following eight variables over a seven race history (two of Chen et. al.'s original 10 variables could not be used because they are specific to greyhound race grades which have no equivalent in harness racing):

♦ Fastest Time – horse's time in the previous race

♦ Win Percentage – the number of wins divided by the number of races

♦ Place Percentage – the number of places divided by the number of races

♦ Show Percentage – the number of shows divided by the number of races

♦ Break Average – the horse's average position out of the starting gate

♦ Finish Average – the horse's average finishing position

♦ Time7 Average – the average finishing time over the last 7 races

♦ Time3 Average – the average finishing time over the last 3 races

5.1 Machine Learning

For our machine learning component we used Support Vector Machines (SVM). SVM is a machine learning classifier that attempts to maximally separate the classes by computing a hyperplane equidistant from the edges of each class [26], as shown in Figure 2.
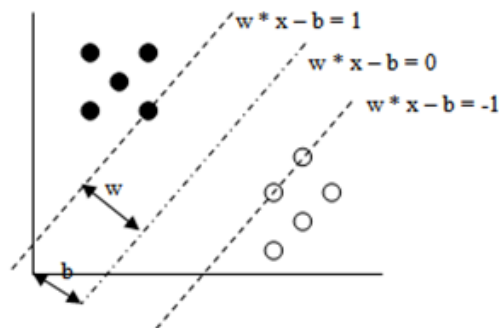

Figure 2. Support Vector Machine

Aside from dividing classes, the hyperplane can also be used as a regression estimate where independent variables are projected on to the hyperplane in order to derive the dependent variable. This variation of SVM is called Support Vector Regression (SVR) which allows for a

continuous numeric prediction instead of classification. This allows us the freedom to rank predicted finishes rather than perform a win/lose classification.

As an example of how the system works once it has been trained, each horse in the testing set is given a predicted finish position by the SVR algorithm. Looking at the horse *Miss HKB* on Oct. 8, 2010 in race 4, we compute the variables for the prior seven races as shown in Table 1.

Fastest Time	119.56
Win Percentage	14.29%
Place Percentage	42.86%
Show Percentage	14.29%
Break Average	5.14
Finish Average	2.71
Time7 Average	120.19
Time3 Average	119.90
Predicted Finish	**2.9259**

Table 1. Miss HKB data

For this particular race, *Miss HKB* has a Fastest Time value of 119.56 which is the time in the last race. *Miss HKB* has a Win percentage of 14.29%, a Place percentage of 42.86% and a Show percentage of 14.29% indicating how often they finish in each of those positions. The Break Average is 5.14 meaning that this horse is typically in the middle to the back of the pack coming out of the gate. The Finish Average is 2.71 indicating plenty of second and third place finishes. The Time7 Average shows that over the past seven races, *Miss HKB* has had an average time of 120.19 seconds. The impressive Time3 Average of 119.90 indicates speed improvement in the last three races. S&C Racing predicts from its internal model that *Miss HKB* should finish 2.9259 which is a good finish, but cannot be fully interpreted until compared with the predicted finishes of other horses in the race. The lower the predicted finish number, the stronger the horse is expected to be and the predicted finish value is independent of the other

horses in the race. As a visual aid, Table 2 shows the race output for Northfield Park's race 4 on October 8, 2010.

Horse Name	Predicted Finish
Miss HKB	2.9259
B B Big Girl	3.9036
Friendly Kathy	4.3016
ShadyPlace	5.2731
St Jated Strike	5.7243
Honey Creek Abby	5.7831
Mad Cap	5.9335
Pamela Lou	6.5066
Winning Yankee	6.8589

Table 2. Predicted Values

From this table, we can wager on *Miss HKB* to win, *B B Big Girl* to place and *Friendly Kathy* to show based on S&C Racing's predicted finish.

5.2 S&C Betting Engine

Once we have the predicted finishes, we pass operations to the Betting Engine to make the appropriate wagers. In order to interrogate the Betting Engine and avoid having it wager on every race, we performed a sensitivity analysis by introducing a variable Cutoff that ranges from 1 to 8 and is incremented by 0.1. The Betting Engine is then limited to wagering on races in which the strongest predicted finish exceeds the Cutoff. Given our previous example of Table 2, if the predicted finish cutoff variable was 2.9, no wagers would be made as *Miss HKB* has a higher predicted finish of 2.9259. If we increment the cutoff to 3.0, now *Miss HKB* is under the cutoff and this race can be wagered upon. Measures are then taken of prediction accuracy, payout and betting efficiency of each wager for each increment. We feel that this type of analysis provides a clearer understanding of performance data than a single value could.

5.3 Crowdsourcing, Dr. Z and Random Chance

In order to best compare the S&C Racing results, we compare them against established prediction methods of crowdsourcing, Dr. Z bettors and random chance. For the crowdsourcing comparison we use pre-race odds where the crowd favorite, the animal with the lowest odds, was selected and wagered upon. For Dr. Z bettors, Place wagers (betting the animal comes in 2nd) were placed on animals with win percentage to place percentage ratios greater than or equal to 1.15, and Show wagers (betting the animal comes in 3rd) were placed with win percentage to show percentage ratios greater than or equal to 1.15. For random chance, a random animal is selected and wagered upon for each race.

6. Experimental Findings and Discussion

6.1 Evaluating a seven race history

To answer our first research question of *can a machine learner predict harness races better than established prediction methods* we present Figure 3a that looks at accuracy through Chen et. al. variables (7 race history) for the three wager types and compares it to established prediction methods in Table 3.
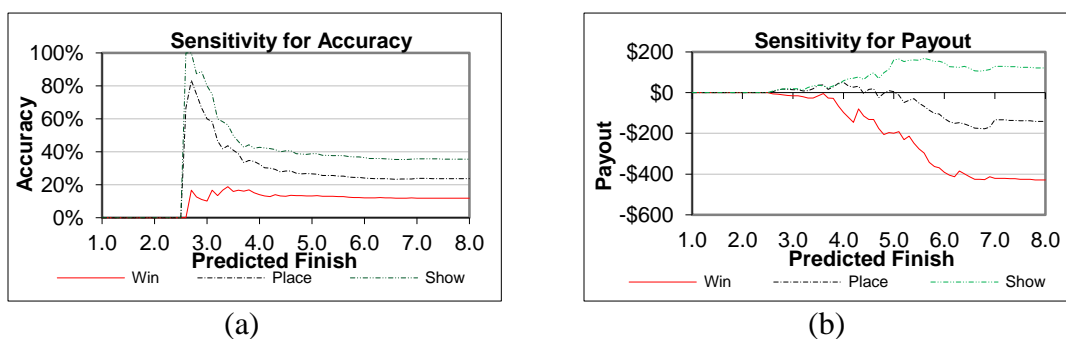

(a)   (b)

Figure 3. S&C Racing for Traditional Wagers (7 race history)

	Accuracy			Payout		
	Win	Place	Show	Win	Place	Show
S&C Racing	18.75% (3.4)	83.33% (2.7)	100.0% (2.7)	-$4.80 (3.6)	$50.70 (4.0)	$168.60 (5.6)
Crowdsourcing	41.67%	65.63%	73.96%	$46.60	$115.00	$119.90
Dr. Z Bettors		24.43%	36.59%		-$44.10	$72.60
Random Chance	11.82%	23.64%	35.45%	-$431.20	-$143.60	$119.50

Table 3. Comparing Established Methods in Traditional Wagers (7 race history)

From our data, Win maxed out at 18.75% at a predicted finish of 3.4 or better, Place maxed out at 83.33% at a predicted finish of 2.7 or better and Show maxed out at 100.0% accuracy at a predicted finish of 2.7 or better.

Comparing S&C Racing to the established prediction methods of Crowdsourcing, Dr. Z Bettors and random chance; S&C Racing underperformed Crowdsourcing on Win (18.75% to 41.67%), but outperformed Crowdsourcing Place (83.33% to 65.63%) and Show (100.0% to 73.96%) (p-values < 0.1 and < 0.05 respectively). S&C Racing also outperformed Dr. Z Bettors (p-values < 0.01) and random chance (p-value < 0.1 for Win, < 0.01 for Place and Show).

Looking at just S&C Racing and Crowdsourcing, both methods increased accuracy when wagering on lower placed finishes (i.e.; moving from Win to Place to Show). This was expected behavior because wagering on these finishes will still garner a payout and count towards accuracy totals if the animal finishes at or higher than the finish wagered upon (as evidenced by random chance). However, S&C Racing exhibited a superior curve in predicting accuracy through the lower placed finishes. Our intuition is that Crowdsourcing was much more efficiently optimizing for Win, perhaps through knowledge external to the system, yet was overtaken by S&C's efficiency when it came to non-win traditional wagers which may imply that S&C Racing is taking advantage of an arbitrage opportunity.

While the results are showing that S&C Racing is performing at least as well as other prediction methods in terms of accuracy, Figure 3b looks at how well S&C Racing compares in terms of wagering payout.

From this data, Win performed poorly in terms of payout, and minimized its losses at ($4.80). Place and Show did better with maximized returns of $50.70 and $168.60 respectively. However, by comparison, S&C Racing Win and Place underperformed Crowdsourcing (-$4.80 to $46.60 and $50.70 to $115.00 respectively), while Show outperformed ($168.60 to $119.90, p-value < 0.01). The S&C Racing wagers both outperformed Dr. Z Bettors and random chance (p-values < 0.01 each). Again, returning to just S&C Racing and Crowdsourcing, payouts increase on the progression through the lower placed finishes which was not unexpected. For S&C Racing, Win started with a negative payout but by Show it exhibited a superior payout. This is further evidence that Crowdsourcing is better optimized on Win but is soon overtaken by S&C Racing.

In terms of exotic wagers, Figure 4a shows the accuracy of S&C Racing's exotic wagers in comparison to the established predictors in Table 4.
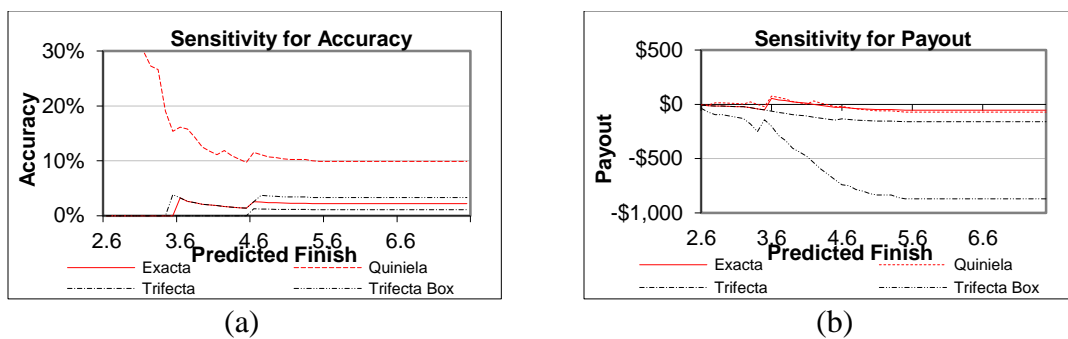

(a)   (b)

Figure 4. S&C Racing for Exotic Wagers (7 race history)

	Accuracy				Payout			
	Exacta	Quiniela	Trifecta	Trifecta Box	Exacta	Quiniela	Trifecta	Trifecta Box
S&C Racing	3.23% (3.6)	37.50% (2.9)	1.28% (4.6)	3.70% (4.7)	$55.60 (3.6)	$76.80 (3.6)	-$6.00	-$36.00
Crowdsourcing	15.38%	26.37%	8.79%	19.78%	$25.40	$56.00	$181.00	$380.40
Random Chance	1.58%	2.79%	0.25%	1.47%	-$1,412.40	-$1,428.60	-$1,517.40	-$4,945.20

Table 4. Comparing Established Methods in Exotic Wagers (7 race history)

From the data, S&C Racing's exotic wagers mostly underperformed Crowdsourcing in terms of accuracy, with the exception of Quiniela, 37.50% to 26.37% (p-value < 0.2). Versus random chance, only the combination wagers of Quiniela and Trifecta Box exhibited a statistically positive difference, 37.50% to 2.79% (p-value < 0.05) and 3.70% to 1.47% (p-value < 0.15) respectively. Although Exacta and Trifecta did exhibit a positive difference versus random chance counterparts, they were statistically equivalent.

Switching to exotic wager payouts, Figure 4b shows us the four S&C Racing wagers on the field of predicted horse strength. From this data, both Exacta and Quiniela outperformed the Crowdsourcing counterparts, $55.60 to $25.40 and $76.80 to $56.00 respectively (p-values < 0.01). The wagers of Trifecta and Trifecta Box underperformed Crowdsourcing which was unexpected given S&C Racing's success at Exacta and Quiniela wagering. S&C Racing also outperformed random chance in all four wager types (p-values < 0.01).

Looking at just S&C Racing and Crowdsourcing between Exacta and Quiniela, again we witness an increase in both Accuracy and Payout which is not unexpected given the better odds of a Quiniela wager versus Exacta. Again S&C Racing exhibits a superior curve between these two wagers versus Crowdsourcing. However, when it came to the Trifecta and Trifecta Box pair, S&C Racing posted dismal results. Taken together, the slow starts and poor performance on Trifecta straight and box, we are reminded that these results are based upon a seven race history handed down from greyhound racing. While the Time7 variable was given the most weight by the SVR algorithm and could be argued to be the most important of the eight input

variables; we contemplate whether this seven race history is leading to calibration issues with

S&C Racing and whether an alternative race history would provide better results.

6.2 Evaluating a four race history

To answer our second research question of *how important is race history to a machine*

learner, we look at maximized accuracy across predictions using race histories of one to ten

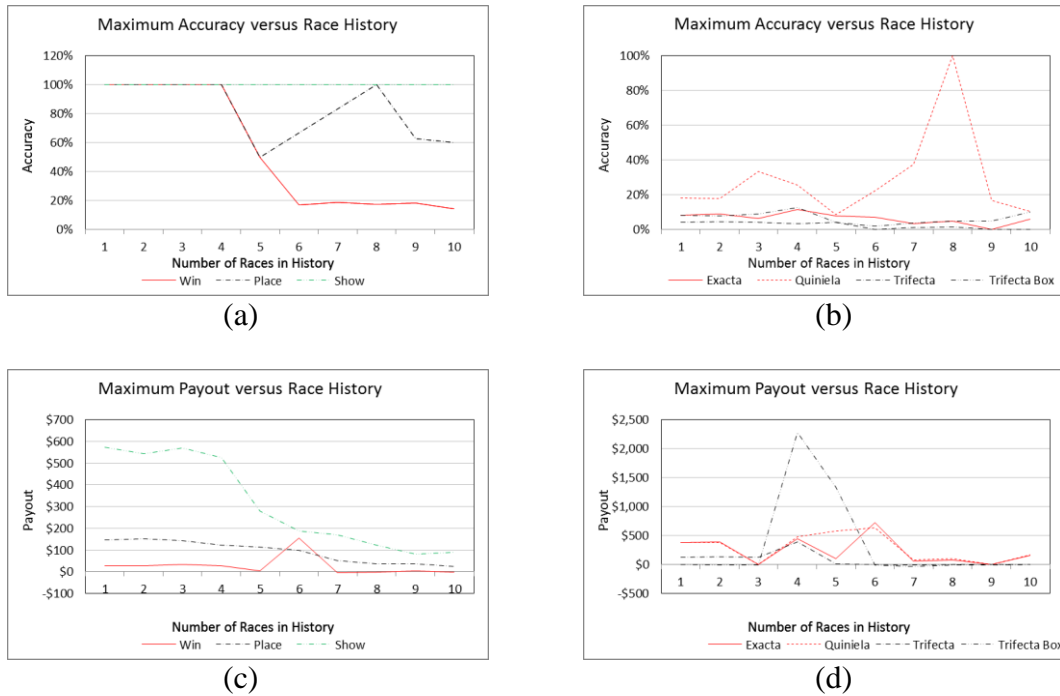races, as shown in Figures 5a and 5b, and maximized payouts as shown in Figures 5c and 5d.

(a)   (b)

(c)   (d)

Figure 5. Examining Wagers Across Race History

Race History	Accuracy	Payout
1	339.03%	$1,621.20
2	339.11%	$1,603.50
3	353.07%	$861.10
4	**353.09%**	**$4,265.40**
5	224.55%	$2,424.60
6	215.03%	$1,791.90
7	247.94%	$304.90
8	328.91%	$325.00
9	202.35%	$108.50
10	200.69%	$438.40

Table 5. Sum of Maximized Accuracies and Payouts Across all Seven Wagers

From these figures, if we were to take all seven wager accuracies together, we would find that the four race history had the best performance (353.09%), albeit marginally, as shown in Table 5. Similarly, by looking at the maximized payouts across race histories as shown in Figures 5c and 5d, we would again find the four race history providing the highest payouts ($4,265.40), Table 5. Taken together, this would indicate that the four race history is performing better than any of the other racing histories, including the seven race history which was used in studies by Chen et. al. (1994) and Schumaker and Johnson (2008). It is interesting to note that the four race history is a trade-off between the intuition of professional wagerers and data miners; whereas the professional wagerer would focus more heavily on the most recent races (following the adage that you are only as good as your last race) and the data miner would want as many races as possible to develop a fuller picture of performance.

We suspect that the four race history could be the result of peculiarities in racing schedule at Northfield Park. There are typically four races over the weekend with a several day break during the midweek. We feel that this midweek break may be introducing confounding variables such as additional training, adjustments to food, medicine and/or routine (e.g.; moving to a different stable or track). Whereas during the weekend race period, less variable change can occur and hence the animals behave in a more predictable manner.

Restating our model in terms of a four race history, there are 698 training races on 5,777 training cases, 194 testing races on 1,653 testing cases with an average of 8.52 horses per race. Looking again at maximizing accuracy of traditional wagers but this time with a four race history, we present Figure 6a and Table 6.
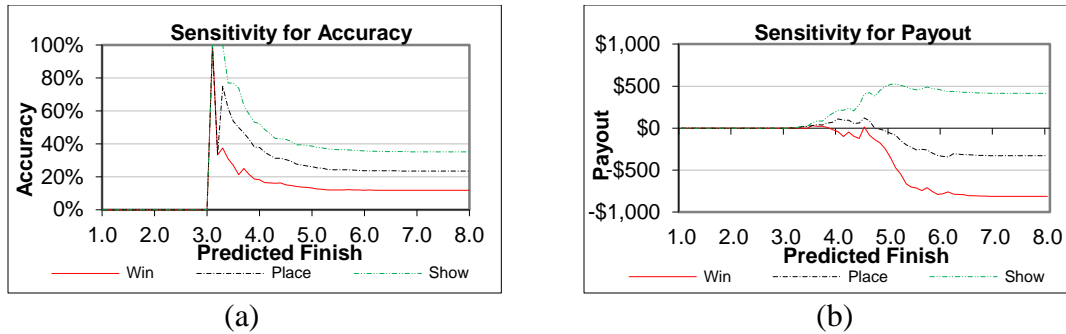
Sensitivity for Accuracy

Accuracy (y-axis): 0%, 20%, 40%, 60%, 80%, 100%
Predicted Finish (x-axis): 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0
Win — Place — Show

(a)

Sensitivity for Payout

Payout (y-axis): -$1,000, -$500, $0, $500, $1,000
Predicted Finish (x-axis): 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0
Win — Place — Show

(b)

Figure 6. S&C Racing for Traditional Wagers (4 race history)

	Accuracy			Payout		
	Win	Place	Show	Win	Place	Show
S&C Racing	37.50% (3.3)	75.00% (3.3)	100.0% (3.3)	$28.20 (3.5)	$122.70 (4.5)	$526.50 (5.1)
Crowdsourcing	45.50%	69.00%	78.00%	$135.40	$257.80	$283.50
Dr. Z Bettors		23.53%	35.21%		$4.30	$103.90
Random Chance	11.74%	23.47%	35.21%	-$813.20	-$328.60	$415.10

Table 6. Comparing Established Methods in Traditional Wagers (4 race history)

From the four race history data, Win still underperformed Crowdsourcing (37.50% to 45.50%), Place was statistically equivalent and Show outperformed its Crowdsourcing counterpart (100.0% to 78.00%) (p-value < 0.01). S&C Racing also outperformed the Dr. Z Bettors (p-values < 0.01) and random chance (p-value < 0.1 for Win, < 0.01 for Place and Show).

Statistically speaking, the accuracies of the four race history exhibited a similar performance to the seven race history seen earlier. However, the differences became apparent when looking at the maximized payouts as shown in Figure 6b. From the data, Win and Place underperformed Crowdsourcing but Show outperformed it $526.50 to $283.50 (p-value < 0.01). S&C Racing also outperformed Dr. Z Bettors and random chance (p-values < 0.01). It is interesting to note the comparatively smaller payout increase in Crowdsourcing versus S&C Racing. This may indicate that Crowdsourcing is experiencing a marginal return on the lower placed finishes (i.e.; Place and Show), whereas S&C Racing is capitalizing on additional knowledge not widely known by the crowds.

In terms of maximized accuracy on exotic wagers, we present Figure 7a and Table 7.
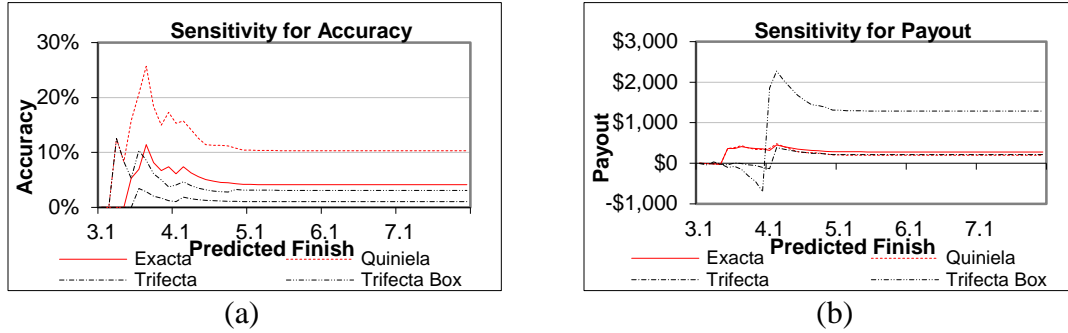

(a)   (b)

Figure 7. S&C Racing for Exotic Wagers (4 race history)

	Accuracy				Payout			
	Exacta	Quiniela	Trifecta	Trifecta Box	Exacta	Quiniela	Trifecta	Trifecta Box
S&C Racing	11.43% (3.7)	25.71% (3.7)	3.45% (3.6)	12.50% (3.3)	$450.40 (4.2)	$479.00 (4.2)	$385.00 (4.2)	$2,273.60 (4.2)
Crowdsourcing	19.59%	30.41%	7.73%	15.98%	$139.40	$197.40	$223.40	-$145.00
Random Chance	1.56%	3.12%	0.24%	1.44%	-$873.60	-$955.00	-$939.00	-$2,170.60

Table 7. Comparing Established Methods in Exotic Wagers (4 race history)

From the data, S&C Racing all underperformed Crowdsourcing in terms of accuracy but outperformed random chance (p-values < 0.01). However, the results were different when looking at payouts in Figure 7b. It can be noted that S&C Racing outperformed all Crowdsourcing and random chance wagers (p-values < 0.01).

Now it becomes a question of calibration. While the seven wager types as a whole had better accuracy than the seven race counterparts, the wagers for Crowdsourcing exhibited slightly better accuracy. However, that accuracy did not translate into better payout gains. By contrast, five of S&C Racing's wagers outperformed in terms of payout leading us to suspect that S&C Racing was able to capitalize on longshot wagers (decreased accuracy, increased payout). We saw a similar accuracy/payout trade-off in Chen et. al. (1994) and Schumaker & Johnson (2008). We feel this result stems from S&C Racing better able to identify the longshot wagers with higher payouts than crowds whom may be wagering more conservatively.

6.3 Evaluating betting efficiency

To answer our third research question of *what wager combinations work best and why*, we analyze each wager in terms of betting efficiency, or the amount of return for every $1 wagered. Figure 8a demonstrates betting efficiency on traditional wagers while Figure 8b shows the efficiencies of exotic wagers. Table 8 provides numerical descriptions of the observed peaks.
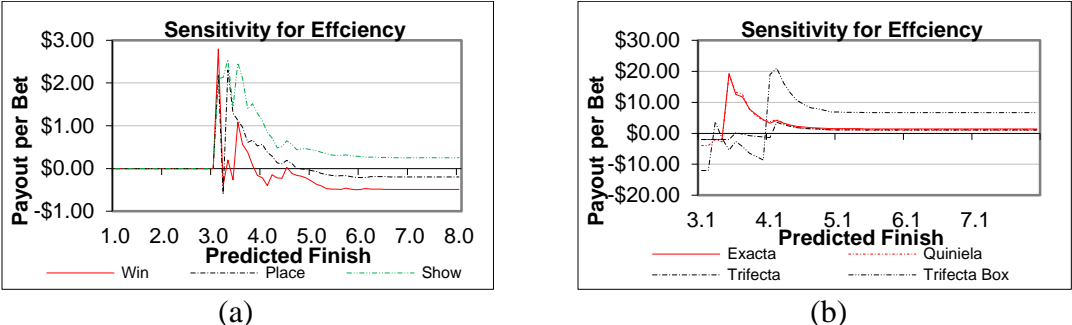

(a)   (b)

Figure 8. Betting Efficiency of Wagers (4 race history)

	Win	Place	Show	Exacta	Quiniela	Trifecta	Trifecta Box
S&C Racing	$1.08 (3.5)	$2.30 (3.3)	$2.55 (3.3)	$19.24 (3.5)	$18.93 (3.5)	$3.56 (4.2)	$21.05 (4.2)
Crowdsourcing	$0.68	$1.29	$1.42	$0.72	$1.02	$1.15	-$0.75
Dr. Z Bettors		$0.00	$0.06				
Random Chance	-$0.49	-$0.20	$0.25	-$4.50	-$4.93	-$4.84	-$11.19

Table 8. Betting Efficiency of Wagers (4 race history)

From the data, S&C Racing outperformed Crowdsourcing in all seven wagers (p-values < 0.01), outperformed the Dr. Z Bettors (p-values < 0.01) and outperformed random chance in all wagers (p-values < 0.01). While it could be argued that S&C Racing is wagering on the strongest races, Dr. Z is similarly selective and Crowdsourcing should be just as strong assuming rational bettors with the same access to information that S&C Racing uses. However, the discrepancies would appear to indicate that S&C Racing is exploiting an informational inequality within the harness racing market. Even more interesting is the Trifecta Box wager; when applied to every race, this wager was still able to obtain a $6.63 return per dollar wagered, which still performs better than Crowdsourcing. Digging further into this result, Trifecta Straight was

26

correct 6 times on 194 wagers (3.09% accuracy), but offset its losses with enough successful

longshot returns.

The other interesting item of note was the performance of the Dr. Z System. It has been

observed that since its widespread usage at tracks, bettors have effectively arbitraged the Dr. Z

System to nil gain. In fact, we observed this effect in Table 8 where Place betting for the Dr. Z

System provides a $0.00 return. If instead of applying the Dr. Z System to every race, as the

strict book implementation would allow, we couple it with a predicted finish measure, like S&C

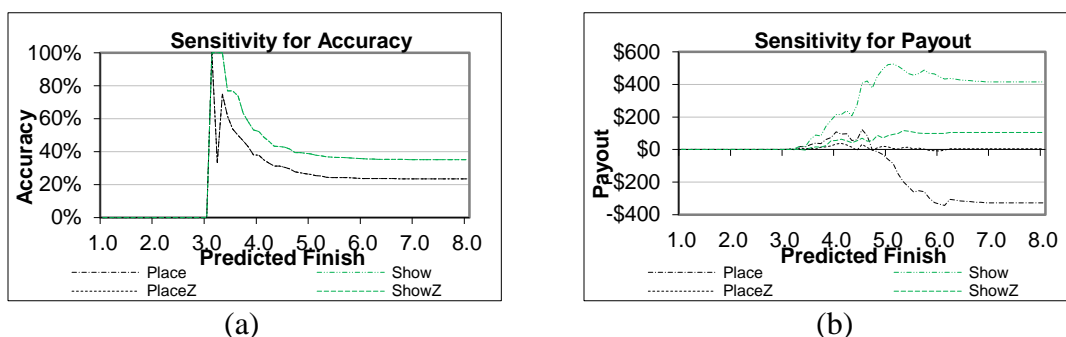Racing, and make wagers accordingly, we would have Figures 9a and 9b.

Figure 9. Comparing Dr. Z to S&C Racing (4 race history)

From Figure 9a, we note that S&C Racing's Place and Dr. Z's PlaceZ run identical, as

does S&C Racing's Show and Dr. Z's ShowZ (running atop one another). This interesting result

indicates that both systems have identical accuracies and hints that S&C Racing may be

employing a similar mechanism as Dr. Z, albeit from a machine learning perspective. However,

if we dig further into this relationship, we notice that the payouts differ in Figure 9b. This means

that while the accuracies were identical; both systems were wagering on different races making

the identical accuracy more of a coincidence. From Figure 9b, S&C Racing's Place

outperformed Dr. Z's PlaceZ until the predicted finish of 4.7 when the roles reversed. For Show,

S&C Racing's Show outperformed Dr. Z's ShowZ on all finishes. It was further interesting to

note that both Show and ShowZ outperformed Place and Place Z wagers, meaning that Show and ShowZ wagers are more profitable.

While the book implementation of Dr. Z has been effectively arbitraged to zero, coupling it with a competent finish strength predictor can provide similar results to S&C Racing in terms of accuracy, but not payout. This observation could perhaps breathe new life into Dr. Z variations.

7. Conclusions and Future Directions

The seven race history has been a defacto standard in several racing studies. We sought to empirically test whether this standard is optimal within the harness industry. By evaluating the seven race history as a baseline, we found that S&C Racing mostly outperformed random chance, Dr. Z bettors and had mixed results versus Crowdsourcing in both accuracy and payouts. Taken alone, S&C Racing shows promise with similar to better performance than the well-established prediction methods.

When evaluating a range of different race histories to track accuracy and payouts, we found that a four race history performed best. Comparing this four race history against established predictors we found that S&C Racing outperformed both random chance and Dr. Z Bettors in both accuracy and payouts. When compared against Crowdsourcing, S&C Racing exhibited a mixture of results with the exception of Show wagers which consistently outperformed Crowdsourcing in both accuracy and payout; and exotic wagers which similarly outperformed the Crowdsourcing counterparts in terms of payouts. We found that S&C Racing was better able to identify top contenders than the crowds were.

In terms of betting efficiency, S&C Racing outperformed all other models easily. It is speculated that S&C Racing was exploiting abnormal payouts arising from informational inequalities within the harness market.

Future directions for this stream of research include tweaking other Chen variables and building a fraud detection framework. This paper investigated manipulating only one of the Chen variables. Perhaps further advances could be found by adjusting other variables as well. For a fraud detection network, now that a robust model of prediction has been built, can we use it to detect unknown/undisclosed injuries or fraudulent behavior (e.g., jockeys/trainers/etc. that consistently over/underperform versus projected finish). Clearly more study within this domain is needed.

References

1.   Progress Publishing, A History of Gambling. Accessed from http://www.crapsdicecontrol.com/gambling_history.htm on Sept. 28, 2011.

2.   Chen, H., et al., 1994. Expert Prediction, Symbolic Learning, and Neural Networks: An Experiment on Greyhound Racing. *IEEE Expert*, 9(6):21-27.

3.   Schumaker, R.P., 2011. Using SVM Regression to Predict Harness Races: A One Year Study on Northfield Park, in *Midwest Decision Sciences Institute Conference*: Indianapolis, IN.

4.   Dana, J. and M. Knetter, 1994. Learning and Efficiency in a Gambling Market. *Management Science*, 40(10):1317-1328.

5.   Surowiecki, J., 2004. *The Wisdom of Crowds* New York: Doubleday.

6.   Pratt, J., 1964. Risk Aversion in the Small and in the Large. *Econometrica*, 32(1-2):122-136.

7.   Hausch, D., W. Ziemba, and M. Rubinstein, 1981. Efficiency of the Market for Racetrack Betting. *Management Science*, 27(12):1435-1452.

8.   Sobel, R. and T. Raines, 2003. An Examination of the Empirical Derivatives of the Favourite-Longshot Bias in Racetrack Betting. *Applied Economics*, 35(4):371-385.

9.   Cain, M., D. Law, and D. Peel, 2003. The Favourite-Longshot Bias, Bookmaker Margins and Insider Trading in a Variety of Betting Markets. *Bulletin of Economic Research*, 55(3):263-273.

10.   Wise, S., 2009. Testing the Effectiveness of Semi-Predictive Markets: Are Fight Fans Smarter than Expert Bookies?, in *Collaborative Innovation Networks Conference*: Savannah, GA.

11.   Spann, M. and B. Skiera, 2008. Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters. *Journal of Forecasting*, 28(1):55-72.

12.   Luckner, S., J. Schroder, and C. Slamka, 2008. On the Forecast Accuracy of Sports Prediction Markets, in Negotiation, Auctions, and Market Engineering, H. Gimpel, et al., Editors, Springer: Berlin. p. 227-234.

13.   Qiu, L., H. Rui, and A. Whinston, 2011. A Twitter-Based Prediction Market: Social Network Approach, in *ICIS*: Shanghai, China.

14.   Hausch, D., V. Lo, and W. Ziemba, 2008. *Efficiency of Racetrack Betting Markets* Singapore: B&JO Enterprises.

15.   Harville, D., 1973. Assigning Probabilities to the Outcomes of Multi-Entry Competitions. *Journal of the American Statistical Association*, 68(342):312-316.

16.   Sauer, R., 1998. The Economics of Wagering Markets. *Journal of Economic Literature*, 36(4):2021-2064.

17.   Ziemba, W. and D. Hausch, 1984. *Beat the Racetrack* San Diego: Harcourt, Brace & Jovanovich.

18.   Ritter, J., 1994. Racetrack Betting - An Example of a Market with Efficient Arbitrage, in Efficiency of Racetrack Betting Markets, D. Hausch, V. Lo, and W. Ziemba, Editors, Academic Press: San Diego.

19.   Tversky, A. and T. Gilovich, 2005. The Cold Facts About the "Hot Hand" in Basketball, in Anthology of Statistics in Sports, J. Albert, J. Bennett, and J. Cochran, Editors, SIAM: Philadelphia.

20.   Albert, J., 2008. Streaky Hitting in Baseball. *Journal of Quantitative Analysis in Sports*, 4(1).

21.   Piatetsky-Shapiro, G., Difference between Data Mining and Statistics. Accessed from http://www.kdnuggets.com/faq/difference-data-mining-statistics.html on Oct 2, 2008.

22.   Schumaker, R., O. Solieman, and H. Chen, 2010. *Sports Data Mining* New York: Springer.

23.   Chen, H. and M. Chau, 2004. Web Mining: Machine Learning for Web Applications. *Annual Review of Information Science and Technology (ARIST)*, 38:289-329.

24.   DataSoftSystems, Data Mining - History and Influences. Accessed from http://www.datasoftsystem.com/articles/article-1380.html on Sept. 2, 2009.

25.   Lazar, A., 2004. Income Prediction via Support Vector Machine, in *International Conference on Machine Learning and Applications*: Louisville, KY.

26.   Vapnik, V., 1995. *The Nature of Statistical Learning Theory* New York: Springer.

27.   Schumaker, R.P., et al., 2012. Evaluating Sentiment in Financial News Articles. *Decision Support Systems*, 53(3):458-464.

28.   Huang, Z., et al., 2004. Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparitive Study. *Decision Support Systems*, 37(4):543-558.

29.   Johansson, U. and C. Sonstrod, 2003. Neural Networks Mine for Gold at the Greyhound Track, in *International Joint Conference on Neural Networks*: Portland, OR.

30. Schumaker, R.P. and J.W. Johnson, 2008. Using SVM Regression to Predict Greyhound Races, in *International Information Management Association (IIMA) Conference*: San Diego, CA.

31. Williams, J. and Y. Li, 2008. A Case Study Using Neural Network Algorithms: Horse Racing Predictions in Jamaica, in *International Conference on Artificial Intelligence*: Las Vegas, NV.