An Investigation of SVM Regression to Predict Longshot Greyhound Races

Robert P. Schumaker[1] and James W. Johnson[2]

[1]Hagan School of Business, Information Systems Department
Iona College, New Rochelle, New York 10801, USA
rschumaker@iona.edu

[2]Director of Information Technology, College of Education
Indiana State University, Terre Haute, Indiana 47809, USA
jimjohnson@indstate.edu

Word Count: 5,643

Abstract

In this paper we investigate the role of machine learning within the domain of Greyhound Racing. We test a Support Vector Regression (SVR) algorithm on 1,953 races across 31 different dog tracks and explore the role of a simple betting engine on a wide range of wager types. From this we triangulated our results on three dimensions of evaluation: accuracy, payout and betting efficiency. We found that accuracy and payouts were inversely linked, where our system could correctly predict Wins 45.35% of the time with a betting efficiency of 87.4% (return per bet) for high accuracy low payout, or predict Superfecta Box wagers with 6.45% accuracy and a 2,195.5% return per bet, corresponding to low accuracy high payout. This implied that AZGreyhound was able to correctly identify longshot dogs and we investigate the reasons why as well as the system's performance.

Keywords: Knowledge Management, Data Mining, Support Vector Regression, Greyhounds

1. Introduction

The ability to predict future events with a certain level of accuracy has its appeal with gamblers and academics alike. This diverse demographic subset seeks to find an edge in predictive sciences, albeit with different motivations. The underlying problem with prediction lies within the problem dynamics, where important parameters are difficult to identify, are constantly shifting and the full effect of selected parameters has not been fully explored. The ultimate question becomes, can profitable predictions be made from the parameters selected?

Greyhound racing is recognized as one of the nation's largest spectator sports. According to the American Greyhound Track Operators Association, it is currently legal in 16 states: Alabama, Arizona, Arkansas, Colorado, Connecticut, Florida, Iowa, Kansas, Massachusetts, New Hampshire, Oregon, Rhode Island, South Dakota, Texas, West Virginia and Wisconsin. Other states do not hold live racing but offer simulcasts (broadcasts of remote races) for betting.

Greyhound racing has a following that parallels horse racing. Many of the same elements in animal athleticism exist in both sports and also the betting on these races. The key to betting is determining how to systemically predict the winners and combinations of winning bets. Bettors must carefully read all the information on the race card and gather as much information about the dogs as possible. Bettors will examine the dogs and their physical conditions, how they have shown in past races, their breeding and bloodlines, their assigned grades for how well they perform, as well as their odds against the dogs they will race against. Weather also plays a role as some bettors rely on the inside traps (part of the track) during wet races. Even weights of the dogs come into play. Lighter dogs and the longshots tend to get bumped and pushed out of the running in the first couple of turns.

Greyhound racing is considered by many to be the most consistent and predictable form of racing (GRAA, 2008). Some expert greyhound bettors recommend keeping bets simple when first starting. Bets on win, place or show are easiest. A win means your dog finished first, place means it finished first or second and show means your dog finished first, second or third. If you can pick the winning dogs in certain categories over a series of races, that will earn more money.

Our research motivation is to build upon prior machine learning techniques in the domain of Greyhound Racing and more closely examine the effect of longshots on racing prediction and payouts. We further examine the impact of non-traditional wager types and more robustly examine their resulting accuracy, payouts and return on wagers.

2. Literature Review

Predicting race outcomes can generally be broken into three distinct areas: mathematical, psychological and data mining techniques. Mathematical areas include Harville formulas (Harville, 1973) which are a collection of equations that establish a rank order of finish by using combinations of joint probabilities (Sauer, 1998). Other mathematical formulas include the Dr. Z System where a potential gambler waits until 2 minutes before race, select those dogs with low odds and bet Place (i.e., the dog will finish in 2nd place or better) on those with a win frequency to place frequency greater than or equal to 1.15 and bet Show (i.e., the dog will finish in 3rd place or better) on those with a win to show frequency greater than or equal to 1.15 (Ziemba and Hausch, 1984). This successful system received considerable attention from both academics and gamblers alike. Subsequent studies later found that bettors were effectively arbitraging the tracks and that any opportunity to capitalize on this system was lost (Ritter, 1994).

In Psychological methods, perhaps the best known method for selecting a dog is the longshot bias. Arrow-Pratt theory suggests that bettors will take on more risk in order to offset their losses (Pratt, 1964; Arrow, 1965). 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 (Hausch, Ziemba and Rubinstein, 1981). Another point of view is to place betting in terms of Stock Market Efficiency. In this approach it is argued that betting on favorites should be as profitable as betting on longshots (Sobel and Raines, 2003). However, this is not the case which leads to a bias towards longshot odds. These types of biases were also found to be prevalent in boxing, cricket, horse racing, snooker and tennis (Cain, Law and Peel, 2003).

In Data Mining methods, simulations can be used to predict outcomes. This method has been used with some degree of success in yacht racing (Philpott, Henderson and Teirney, 2004) and the thoroughbred industry (Burns, Enns and Garrick, 2006). However, simulated data does not address the complexities involved with large numbers of varying parameters. Another technique in Data Mining is to use Statistical Learning methods. These systems are better able to generalize the data into recognizable patterns (Lazar, 2004). One of the more recent methods in statistical learning is Support Vector Regression (SVR) which is a variant of Support Vector Machines (SVM) (Vapnik, 1995). Both SVM and SVR are at their essence classification algorithms that seek to maximally classify high dimension data while minimizing their fitting error. SVR differs in the respect that the hyperplane used to divide the sets can be used as a regression estimator and can return discrete values instead of categories. This technique was used in a similar context to predict stock prices from financial news articles (Schumaker and Chen, 2008).

In a prior study of greyhound races, Chen et. al. tested an ID3 and Back Propagation Neural Network (BPNN) on ten race performance-related variables determined by domain experts on 100 races at Tucson Greyhound Park (Chen, Rinde, She, Sutjahjo, Sommer and Neely, 1994). From their work they made binary decisions as to whether the greyhound would finish first based on its historical race data. If a dog was predicted to finish first, they would make a $2 wager. The ID3 algorithm resulted in 34% accuracy and a $69.20 payout while the BPNN had 20% accuracy and a $124.80 payout. This seeming disparity in decreased accuracy and increased payout is justified with the argument that the BPNN was more successful in selecting longshot winners, hence accuracy would suffer but the long odds would result in the higher payouts. By comparing their machine learning techniques to track experts, the experts managed a much more disappointing 18% accuracy and a payout loss of $67.60.

In a follow-up study that expanded the number of variables studied to 18, Johansson and Sonstrod used a similar 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) (Johansson and Sonstrod, 2003). 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 better accuracy and worse payout than Chen et. al. would imply that it was either the additional variables or too few training cases (449 as compared to Chen's 1,600) that harmed their ability to capture longshots. However, their exotic wagers did better. Quiniela had 8.8% accuracy and a $20.30 payout, while Exacta had 6.1% accuracy and $114.10 payout.

The drawbacks of the BPNN design is that there is a binary assignment of win or lose as shown in Table 1.

Race #	Dog # - Name	BPNN
1	4 - Kma Baklava	1
1	2 - Coldwater Bravo	1
1	7 - Dollar Fa Dollar	1
1	3 - Stat U S Mistic	0
1	8 - Bf Oxbow Tiger	0
1	6 - Flyin Low	0
1	5 - Jr B-s Diesel	0
1	1 - Shining Dragon	0

Table 1. BPNN Binary Assignment

In Table 1, three dogs had strong enough historical data to warrant the system to pick them to finish first. This is because each dog's chance of winning is evaluated independently of others. In this situation, you would bet $2 of each of the three to win, guaranteeing at least a $4 loss. This arrangement also does not provide a rank ordering of finish. Is "Kma Baklava" more likely to finish in first place than "Coldwater Bravo?" It also begs the question of how to bet on Place, Show, Exacta, etc.

From our investigation, we found several research gaps. The first of which is that prior studies in Greyhound racing have mostly relied on neural networks. Perhaps by using newer statistical machine learning techniques, better results can be obtained. Also, accuracy and payout was found to be static measures. The techniques used binary decision making which impacted the experiments and as a result there is only one measurement of accuracy and payout. Perhaps by using a strata of different cutoffs between winner and loser, we can fine-tune the tradeoff between accuracy and payouts. These higher cutoffs (i.e., betting on the more longshot dogs) may lead to higher payouts whereas lower cutoffs (i.e., betting on the strongest dogs) may lead to higher accuracy.

3. Research Questions

From our analysis we propose the following research questions.

- *How accurate is a machine learning method in predicting Greyhound race outcomes?*

We believe that using an SVR style approach may lead to better results than BPNN. SVR also has the added bonus of being able to return discrete values rather than strict classifications which can allow for a deeper inspection of accuracy results.

- *How profitable is the same system?*

From prior research we know that accuracy and payout are inversely related. By increasing the amount of wagering risk to include longshots, we believe that there exists an optimal arrangement where payout can be maximized.

- *How will the addition of exotic wagers affect system accuracy and profitability?*

The Johansson and Sonstrod study began the investigation of exotic wagering through Quiniela and Exactas. Perhaps investigating other forms of exotic wagering could lead to better insight into longshot bias.

4. System Design

To address these research questions, we built the AZGreyhound system as shown in Figure 1.

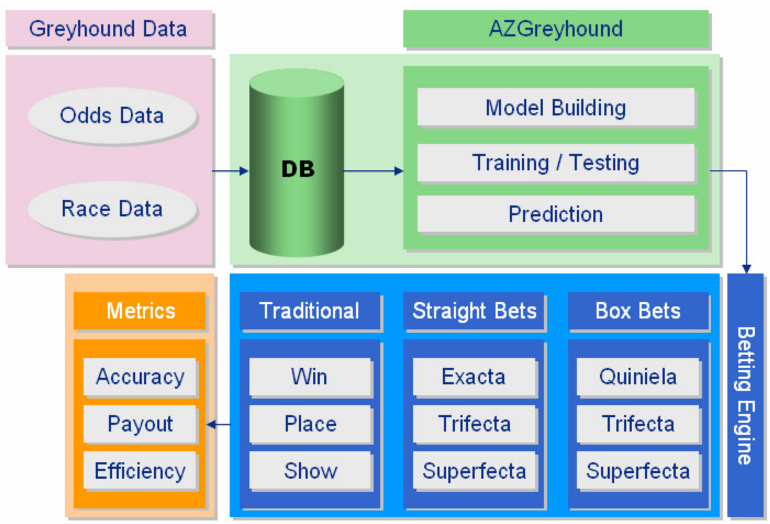

Figure 1. The AZGreyhound System

The AZGreyhound system consists of several major components: the data gathering module, the machine learning part, the betting engine and the evaluation metrics. The odds data is the individual race odds for each wager type (e.g., Win, Place, Show, etc.). The race data are features gathered from the race program. A sample race program is shown in Figure 2.

```
4TH RACE  Tucson - 03/05/07E Grade: D, Distance: 5/16
1st Leg $2 Straight $1 Wheel Pick 4 / Wps, Quiniela, Exacta, Trifecta & 10cent Superfecta Wagering

1  Gt's Red Bear            (55)  31.69  Starts: TU  15  1  2  2  1
Breedline: Red, F, 09/09/04, Sire: Bl's Browny Bear, Dam: Go Sharon Go
Kennel: Cayla Young, Trainer: Young, Cayla
02/28/07E 5 TU 5/16 F 30.94 54.0  7 7 3   3   2-4.5    31.27  3.60  TD  Drove For Place, Md  PhtAzJssc,BbUnrlyFvr,RdBgRd7
02/24/07E 2 TU 5/16 F 31.30 55.5  6 6 3   2   2-1      31.38  8.40  D   Not Far Off, Wide    LphRalbrd,TnjChlDg,KwCbrn7
02/21/07E 5 TU 5/16 F 31.19 55.5  8 8 8   8   7-10.5   31.91  9.40  D   Not In It            GlbsStrmy,Shsttsprs,PhnxGm8
02/17/07E14 TU 5/16 F 30.55 54.0  7 4 5   5   6-11.5   31.37 45.40  C   Gradual Decline      CmThndr,FgscnchSss,BttysGl7
02/14/07E 5 TU 5/16 F 30.84 55.0  5 5 6   6   7-11.5   31.64 10.70  C   Slight Fade          IslandPharoh,UlsJJ,SkySng8
2  Rebel Annie              (64)  31.12  Starts: TU  14  1  1  2  2
Breedline: Black, F, 03/06/03, Sire: Nebraska Rebel, Dam: Turbo Strategic
Kennel: Curtis L Washburn, Trainer: Ross, Robert
03/01/07E 2 TU 5/16 F 30.85 65.0  2 5 7   7   4-7.5    31.39  7.90  D   Bumped 1st Turn      DtchShndr,UssNwsFlsh,CldwtrRnls8
02/26/07E 3 TU 5/16 F 31.10 65.0  2 4 4   5   4-5.5    31.50 49.00  C   No Mishaps           KmaEyeSeeEwe,ByJsmn,PpJnn7
02/22/07E 9 TU 5/16 F 30.90 64.5  3 4 5   5   7-12.5   31.79  8.80  C   Faded Back           VLxstrr,BlckStrlt,EmprSttsmn7
02/19/07E 4 TU 5/16 F 31.06 65.0  6 7 5   5   5-6.5    31.50  6.90  C   Blocked 1st Turn     PhoenxPkr,DlyDty,KmDynsty7
02/15/07E 2 TU 5/16 F 30.85 65.0  7 7 6   8   7-8      31.40 49.00  C   Not In It            RjsCttlKt,WthABng,PtCFrntLn8
```

Figure 2. A sample race program

Each race program contains a wealth of data. There are generally 12 races per program where each race has 6 to 8 dogs. The usual number of dogs per race is eight, but some dogs may scratch (i.e., not race) which can lower the field of competition. Also within the race program, each dog has the results from the previous 5 races. There is some dog-specific data within the race program such as the dog's name, color, gender, birthday, sire, dam, trainer and kennel. Race-specific race information includes the race date, track, fastest time, break position, eighth-mile position, far turn position, finish position, lengths won or lost by, average run time, grade of race, track condition and racing weight.

Once the system has been trained on the data provided, the results are tested along three dimensions of evaluation: accuracy, payout and efficiency. Accuracy is simply the number of winning bets divided by the number of bets made. Payout is the monetary gain or loss derived from the wager. Efficiency is the payout divided by the number of bets which is used for comparative purposes to the prior studies.

The Betting Engine examines three different types of wagers: traditional, straight bets and box bets. In traditional wagers, bettors speculate on whether a dog will win, place or show. If betting on a Win, the bettor receives a payout only if the selected dog comes in first place. If betting on Place, the bettor receives differing payouts if the selected dog comes in either first or second place. If betting on Show, the bettor receives differing payouts if the selected dog comes in first, second or third place. In straight bets, bettors consider the finish placement of multiple dogs through Exacta, Trifecta and Superfecta wagers. In Exacta, the bettor is trying to predict which two dogs will come in 1st and 2nd place respectively. In Trifecta, the job is made more difficult by trying to guess the placement of the first 3 dogs in order. Superfecta is even more difficult where the bettor is trying to determine which four dogs will cross the finish line in order. Box bets simplify the selection process by taking finish order out of the equation. In essence you are betting on every combination of finish between the selected dogs. This makes box betting a more expensive wager.

5. Experimental Design

To perform our experiment, we automatically gathered data from www.trackinfo.com which consists of daily race programs and odds charts for all US Greyhound tracks in operation at the time of this study. Some tracks contained multiple daily programs incorporating both afternoon and evening racing. We eliminated schooling races from the data, as they simply assign race grades to greyhounds and do not contain the full amount of race data. Once the data was gathered, it is parsed to obtain specific race data and sent to AZGreyhound for prediction.

For our collection of greyhound races we chose a study period of January 7 through March 7, 2007. Prior studies used only one racetrack, input their data manually and had small datasets. Chen et. al. (1994) used 1600 training cases from Tucson Greyhound Park, whereas Johansson and Sonstrod (2003) used 449 training case from Gulf Greyhound Park in Texas. Our study differs by automatically gathering race data from multiple tracks. In all, we gathered 41,473 training cases covering 7,760 races. However, we were only able to use 1,953 races because race programs list the race results of the prior five races and we needed data on the prior seven races. This data incorporated 7,163 dogs from 31 different tracks, however, 7 of the tracks provided the bulk of data as shown in Table 2.

Track	# Races
Caliente	1408
Raynham/Taunton	1323
Lincoln	1174
Wichita	962
Melbounre	898
Tucson	788
Hinsdale	700
All Others	507

Table 2. Number of races gathered from various tracks

Using Chen et. al. (1994) as a guide, we limited ourselves to 10 race variables over the most recent seven races for each greyhound:

- Fastest Time – a scaled difference between the time of the race winner and the dog in question, where slower dogs experience larger positive values.
- 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 dog's average position out of the starting box
- Finish Average – the dog'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

- Grade Average – the average racing grade of the dog (A-D) of the last 7 races
- UpGrade – additional points given to a dog racing in a less competitive grade (e.g., +3 points if the most recent race was a better grade, +2 points if the drop in grade was 2 races back, +1 if the drop was 3 races ago, 0 otherwise)

As an example of how the system works, each dog in each race is given a predicted finish position by the SVR algorithm. Looking at Rebel Annie from Figure 2, we compute the 10 variables for the prior seven races as shown in Figure 3.

Dog Name	Track	Race Date	Race #	Fastest Time	Win Percentage	Place Percentage	Show Percentage	Break Average	Finish Average	Time7 Average	Time3 Average	Grade Average	UpGrade	Predicted Finish
Rebel Annie	Tucson	3/5/2007	4	0.0	6.67%	6.67%	13.33%	4.27	4.27	0.43	0.61	1.47	2	1.7604

Figure 3. Rebel Annie variable data

The data is fed into the SVR algorithm which then determines the expected finish position of the dog, 1.7604, which means that the dog is expected to finish between 1st and 2nd place in an 8 dog race. The lower the predicted finish number, the stronger the dog is expected to be. The predicted finish value is independent of the other dogs in the race. As a visual aid, Table 3 shows the race output for Tucson Greyhound Park for Race 1 on Mar 28, 2007.

Race #	Dog # - Name	Predicted Finish
1	4 - Kma Baklava	1.9805
1	2 - Coldwater Bravo	4.2280
1	7 - Dollar Fa Dollar	4.5229
1	3 - Stat U S Mistic	5.0165
1	8 - Bf Oxbow Tiger	5.0811
1	6 - Flyin Low	5.4416
1	5 - Jr B-s Diesel	5.7437
1	1 - Shining Dragon	7.0226

Table 3. Predicted Values for Race 1 on Mar. 28, 2007 at Tucson Greyhound Park

From Table 3, *Kma Baklava* is predicted to win, *Coldwater Bravo* will place and *Dollar Fa Dollar* will show. For the traditional wagers of win, place and show we adopted the same betting engine used in Chen et. al. (1994) where bets are made on individual greyhounds. In this engine, if two greyhounds appear strong within a race, the betting engine will bet on both greyhounds to win. We follow the same approach with place and show as well. The pseudo-code for the betting engine algorithm is as follows:

For each greyhound {

If 'SVR Prediction' <= Arbitrary Cutoff (e.g., 2.0) then {

If 'Finish Position' = 1 then Win else Lose

} }

We vary the *Arbitrary Cutoff* within the experiment between 1 and 8 in 0.1 increments for a sensitivity analysis of AZGreyhound's predictions.

For the betting engine on exotic wagers (e.g., quiniela, exacta, trifecta and superfecta) we adopt the betting engine of Johansson and Sonstrod (2003) and make bets per race instead of per dog, because multiple dogs are predicted to finish in particular orders. The pseudo-code for the exotic wager betting engine (Exacta) is as follows:

For each race, ordered by SVR Prediction ascending {

If (1^{st} SVR Prediction) <= Arbitrary Cutoff then {

If '1^{st} Dog to Finish' = 1 and '2^{nd} Dog to Finish' = 2 then Win else Lose

} }

We chose to use the lowest SVR Prediction of the race for the cutoff value. We also varied the arbitrary cutoff within the experiment between 1 and 8 for a sensitivity analysis.

For the machine learning module we implemented Support Vector Regression (SVR) using the Sequential Minimal Optimization (SMO) (Platt, 1999) function through Weka (Witten and Eibe, 2005). SVR allows for discrete numeric prediction instead of classification. We also selected a linear kernel and used ten-fold cross-validation. This method was used in forecasting futures contracts (Tay and Cao, 2001) and future stock prices (Schumaker and Chen, 2006).

6. Experimental Findings and Discussion

To answer our first research question, on *how accurate is a machine learning method in predicting Greyhound race outcomes*, we performed a sensitivity analysis by varying the cutoff's from 1.0 to 8.0 on Win, Place and Show as shown in Figure 4.

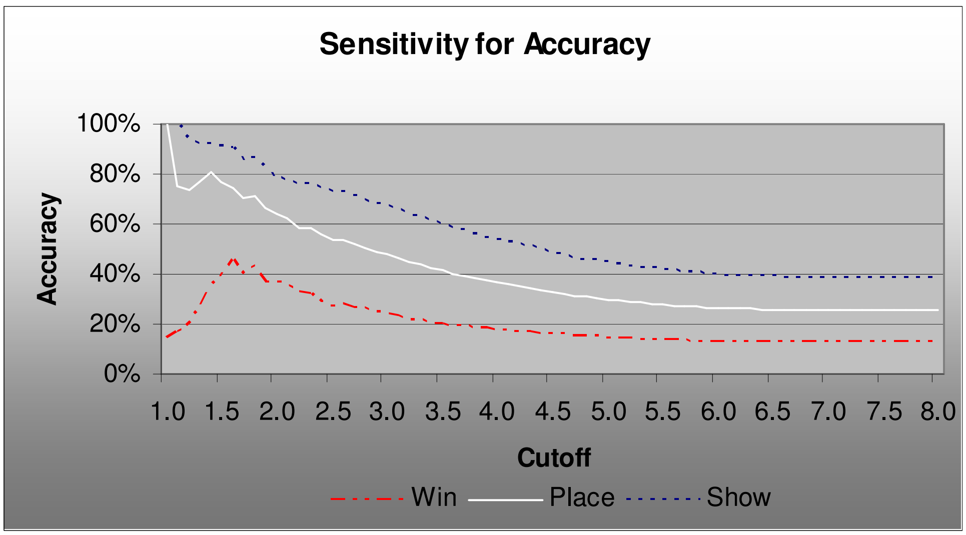

Figure 4. System Accuracy for Traditional Wagers

From Figure 4, Show was most accurate and Win was least accurate as expected because betting Show will provide a positive return if the greyhound Wins, Places or Shows. It is most interesting to note the variation in results that occur between Cutoff 1.0 and 2.5 as shown in Table 4.

	Bet on Win			Bet on Place			Bet on Show		
CutOff	# Correct	# Bets	Accuracy	# Correct	# Bets	Accuracy	# Correct	# Bets	Accuracy
1.0	1	7	14.29%	7	7	**100.00%**	7	7	**100.00%**
1.1	2	12	16.67%	9	12	75.00%	12	12	**100.00%**
1.2	3	15	20.00%	11	15	73.33%	14	15	93.33%
1.3	7	26	26.92%	20	26	76.92%	24	26	92.31%
1.4	13	37	35.14%	30	37	81.08%	34	37	91.89%
1.5	22	56	39.29%	43	56	76.79%	51	56	91.07%
1.6	39	86	**45.35%**	64	86	74.42%	78	86	90.70%
1.7	45	112	40.18%	79	112	70.54%	96	112	85.71%
1.8	60	139	43.17%	99	139	71.22%	120	139	86.33%
1.9	69	186	37.10%	123	186	66.13%	153	186	82.26%
2.0	90	243	37.04%	155	243	63.79%	191	243	78.60%
2.1	118	325	36.31%	203	325	62.46%	252	325	77.54%
2.2	133	406	32.76%	237	406	58.37%	308	406	75.86%
2.3	161	505	31.88%	296	505	58.61%	383	505	75.84%
2.4	187	649	28.81%	364	649	56.09%	481	649	74.11%
2.5	227	825	27.52%	444	825	53.82%	603	825	73.09%

Table 4. System Accuracy for Traditional Wagers

Betting on Win had 45.35% accuracy by betting on greyhounds with a predicted finish of 1.6 or less. Betting on Place had 100% accuracy on 7 instances with a Cutoff of 1.0 and betting on Show was 100% accurate with Cutoffs below 1.2.

Given that there were an average of 7.03 dogs per race, the random probability of selecting a Win wager is 14.22%. From Chen et. al. (1994), their BPNN had 20% accuracy on Wins and Johansson and Sonstrod (2003) had 24.9% accuracy on Wins. AZGreyhound managed statistically better accuracy than random chance with Cutoff's between 1.3 and 5.2 (p-value < 0.01) and peaked at 45.35% accuracy at Cutoff 1.6. Past Cutoff 5.2, where the system is essentially betting on over half of the dogs to Win, AZGreyhound had statistically worse accuracy (p-value < 0.01) as expected.

For Place, the random probability of selecting a winning Place wager is 28.45%. AZGreyhound managed statistically better accuracy than random chance with Cutoff's between 1.0 and 5.3 (p-value < 0.01) and peaked at 100% accuracy at Cutoff 1.0. Past Cutoff 5.3, AZGreyhound had statistically worse accuracy (p-value < 0.01).

For Show, the random probability of selecting a winning Show wager is 42.67%. AZGreyhound managed statistically better accuracy than random chance with Cutoff's between 1.0 and 5.2 (p-value < 0.01) and peaked at 100% accuracy for Cutoffs below 1.2. Past Cutoff 5.2, AZGreyhound had statistically worse accuracy (p-value < 0.01).

To answer our second research question of *how profitable is the same system*, we performed a sensitivity analysis of system payout as shown in Figure 5.

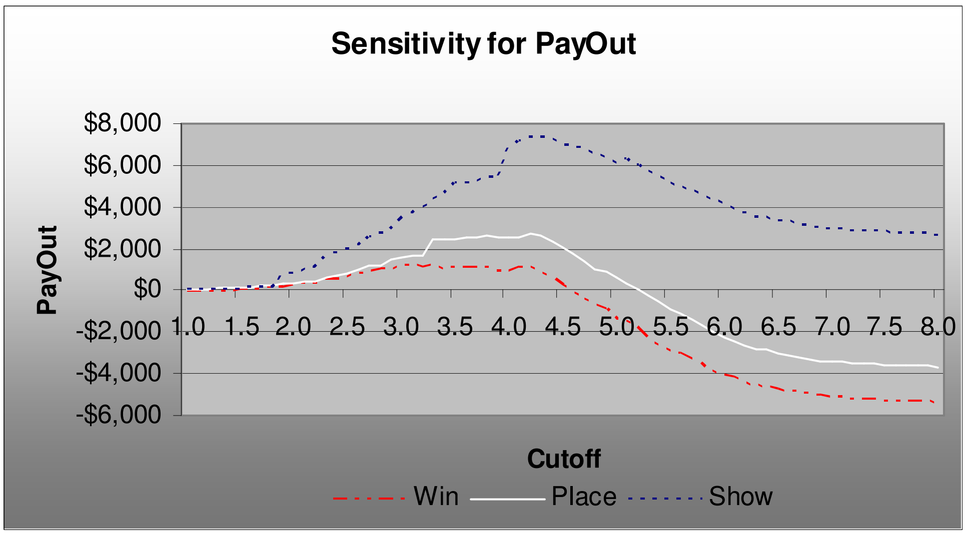

Figure 5. System Payout for Traditional Wagers

From Figure 5, Show had the best payouts and peaked at Cutoff 4.2 with a $7,341.95 payout on 8,289 bets. Win, Place and Show all had positive payouts between cutoffs 1.6 and 4.6. Table 5 looks closer at the cutoff range of 3.5 to 5.0 where all three traditional wagers had the highest payouts.

	Bet on Win			Bet on Place			Bet on Show		
CutOff	# Correct	# Bets	Payout	# Correct	# Bets	Payout	# Correct	# Bets	Payout
3.5	822	4,132	$1,096.10	1,710	4,132	$2,421.37	2,488	4,132	$5,100.61
3.6	914	4,682	$1,102.60	1,889	4,682	$2,524.07	2,739	4,682	$5,110.21
3.7	998	5,220	$1,097.20	2,055	5,220	$2,490.57	2,988	5,220	$5,238.41
3.8	1,076	5,780	$1,065.20	2,223	5,780	$2,602.07	3,240	5,780	$5,434.81
3.9	1,159	6,396	$941.30	2,397	6,396	$2,554.27	3,490	6,396	$5,367.41
4.0	1,228	6,987	$892.10	2,553	6,987	$2,506.37	3,737	6,987	$6,876.45
4.1	1,332	7,650	$1,142.90	2,732	7,650	$2,537.57	4,009	7,650	$7,117.95
4.2	1,421	8,289	$1,086.60	2,918	8,289	**$2,720.17**	4,286	8,289	**$7,341.95**
4.3	1,491	8,901	$811.40	3,068	8,901	$2,631.07	4,525	8,901	$7,307.95
4.4	1,567	9,566	$586.80	3,225	9,566	$2,357.97	4,766	9,566	$7,195.75
4.5	1,640	10,208	$276.30	3,351	10,208	$2,068.97	4,964	10,208	$6,977.95
4.6	1,708	10,825	-$47.70	3,474	10,825	$1,780.77	5,147	10,825	$6,937.15
4.7	1,760	11,382	-$458.10	3,583	11,382	$1,407.67	5,316	11,382	$6,789.25
4.8	1,814	11,915	-$695.30	3,676	11,915	$1,004.37	5,475	11,915	$6,591.65
4.9	1,859	12,446	-$960.60	3,772	12,446	$885.47	5,632	12,446	$6,413.65
5.0	1,901	12,936	-$1,397.40	3,853	12,936	$570.77	5,760	12,936	$6,073.45

Table 5. System Payout for Traditional Wagers

From this table, we can start to see the effect of obtaining larger payouts when betting on the higher cutoffs, i.e., picking the longshots. Win had a maximum payout of $1,248.40 at cutoff 3.1, Place had a maximum payout of $2,720.17 at cutoff 4.2 and Show had a maximum payout of $7,341.95 at cutoff 4.2. Also, the number of bets made is quite large at these higher cutoffs.

In looking at the third metric of betting efficiency on traditional wagers, Figure 6 demonstrates a sensitivity analysis of the differing cutoffs versus the payout per bet.

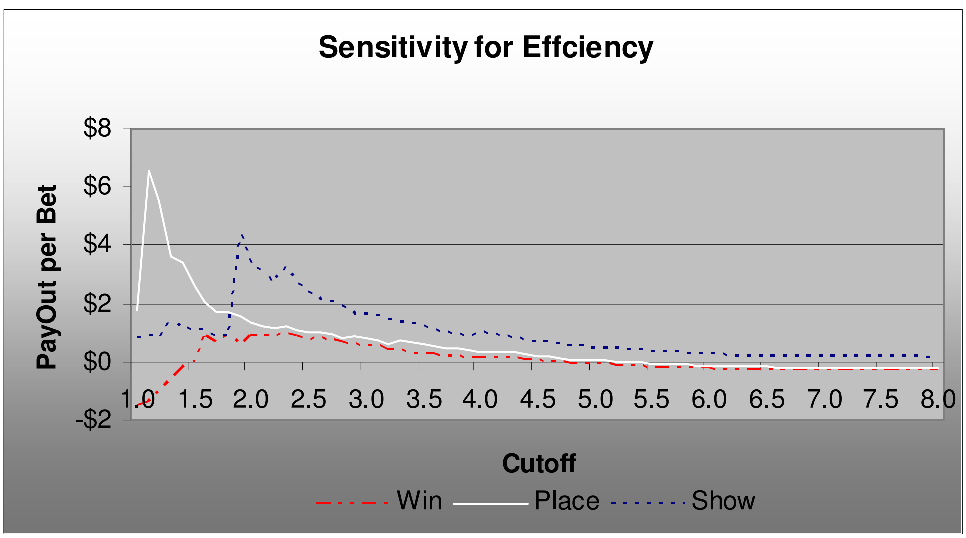

Figure 6. System Efficiency for Traditional Wagers

From Figure 6, Place had the best payout per bet efficiency, peaking at $6.56 return per bet at cutoff 1.1. Win, Place and Show all exhibit positive returns between cutoffs 1.6 and 4.5. The area between cutoffs 1.0 and 2.5 exhibits the most volatility and are shown in detail in Table 6.

	Bet on Win			Bet on Place			Bet on Show		
CutOff	# Correct	# Bets	Efficiency	# Correct	# Bets	Efficiency	# Correct	# Bets	Efficiency
1.0	1	7	-$1.543	7	7	$1.743	7	7	$0.800
1.1	2	12	-$1.383	9	12	**$6.564**	12	12	$0.867
1.2	3	15	-$1.027	11	15	$5.558	14	15	$0.800
1.3	7	26	-$0.638	20	26	$3.622	24	26	$1.392
1.4	13	37	-$0.195	30	37	$3.410	34	37	$1.181
1.5	22	56	-$0.007	43	56	$2.564	51	56	$1.059
1.6	39	86	$0.874	64	86	$2.033	78	86	$1.108
1.7	45	112	$0.638	79	112	$1.708	96	112	$0.891
1.8	60	139	$0.869	99	139	$1.711	120	139	$0.852
1.9	69	186	$0.624	123	186	$1.580	153	186	**$4.370**
2.0	90	243	$0.900	155	243	$1.373	191	243	$3.444
2.1	118	325	$0.892	203	325	$1.205	252	325	$3.070
2.2	133	406	$0.908	237	406	$1.147	308	406	$2.732
2.3	161	505	**$0.946**	296	505	$1.202	383	505	$3.304
2.4	187	649	$0.851	364	649	$1.101	481	649	$2.710
2.5	227	825	$0.709	444	825	$0.992	603	825	$2.358

Table 6. System Efficiency for Traditional Wagers

From this table, Place had the best betting efficiency of $6.564 return per bet at cutoff 1.1. Show was second-most efficient at $4.370 return at cutoff 1.9 and Win was least efficient with its maximum return of $0.946 at cutoff 2.3.

Given the prior studies that predicted greyhound Wins, Chen et. al. (1994) had a payout of $124.80 per 100 bets or a payout efficiency of $1.248 while Johansson and Sonstrod (2003) had a payout loss of -$6.60 per 100 bets or payout efficiency of -$0.066. AZGreyhound's best Win payout was $477.80 per 505 bets or a payout efficiency of $0.946. This means that AZGreyhound did not have as efficient of betting strategy as Chen et. al. (1994). We instead found that we could either have high accuracy and low payout or low accuracy and high payout. This has to do with AZGreyhound being able to successfully predict longshot bets at the higher cutoffs.

To answer our third research question, *how will the addition of exotic wagers affect system accuracy and profitability*, we analyzed the addition of straight and box wagers to our three metrics of accuracy, payout and betting efficiency as shown in Figure 7.

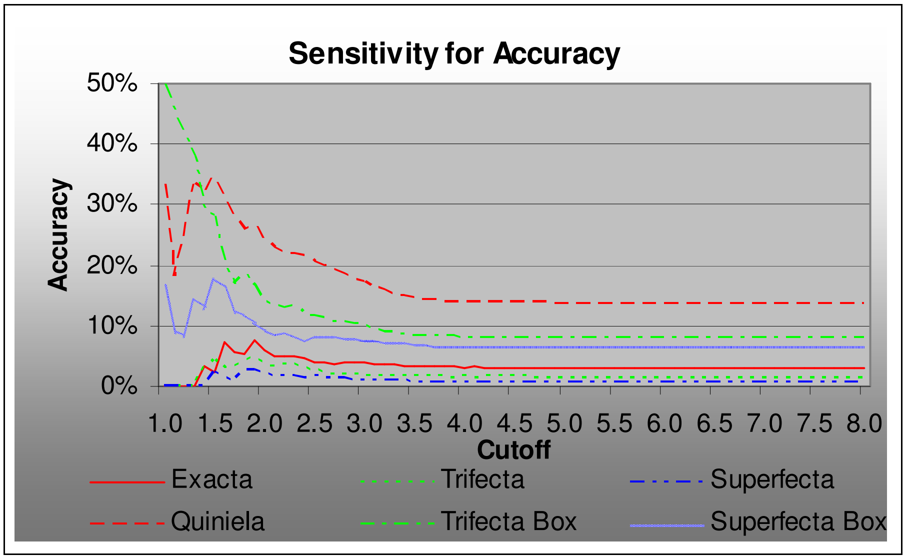

Figure 7. System Accuracy for Exotic Wagers

From Figure 7, all three box wagers, Quiniela, Trifecta Box and Superfecta Box performed better than their straight wager counterparts; Exacta, Trifecta and Superfecta respectively. Trifecta Box had the best accuracy, 50.0% at cutoff 1.0 where it was correct on 3 of 6 bets. Quiniela was second best at 34.8% accuracy at cutoff 1.5. Given the average of 7.03 dogs per race, the random probability of selecting a winning Exacta wager is 2.36%. AZGreyhound managed statistically better accuracy than random chance with cutoff 1.6 and greater, peaking at 7.53% accuracy at cutoff 1.9 (p-value < 0.01). For comparative purposes, Johansson and Sonstrod (2003) had 6.1% accuracy. The random probability of selecting a winning straight trifecta wager is 0.47%. AZGreyhound managed statistically better accuracy than random chance with cutoff 1.8 and greater (p-value < 0.01) and peaked at 4.79% accuracy at cutoff 1.9. Random chance for a straight superfecta wager is 0.12%, however, AZGreyhound obtained statistically better accuracy at cutoff 1.9 and above, peaking at 2.74% accuracy and cutoff 1.9 (p-value < 0.01). For the box wagers, AZGreyhound performed even better. The random probability of selecting a winning quiniela combination is 4.72% and Johansson and Sonstrod (2003) had an impressive 8.8% accuracy. AZGreyhound performed much better with 34.78% peak accuracy at cutoff 1.5 and statistically better accuracy versus random chance, on cutoffs greater than 1.3 (p-value < 0.01). For Trifecta box wagers, the random probability of selecting the winning combination is 2.81%. AZGreyhound peaked at 50.0% accuracy at cutoff 1.0 and managed statistically better accuracy results for cutoffs 1.1 and greater (p-value < 0.01). The probability of correctly selecting a winning Superfecta box wager is 2.79%. AZGreyhound peaked at 17.39% accuracy at cutoff 1.5 and had statistically better accuracy for cutoffs 1.5 and greater (p-value < 0.01).

Looking at our second metric of payout, Figure 8 shows the sensitivity analysis for payout on exotic wagers.

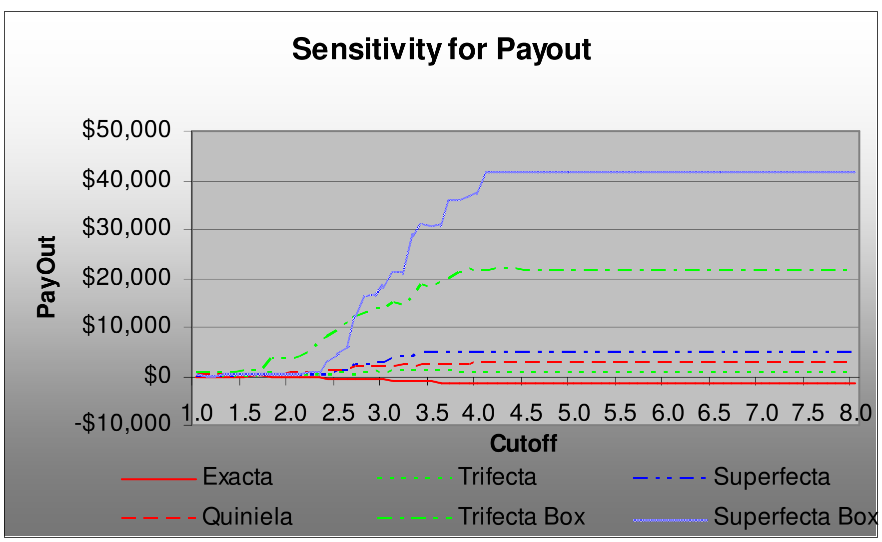

Figure 8. System Payout for Exotic Wagers

From this figure, we can see that Superfecta Box had the highest payout at $41,517.37 at cutoff 4.1 with 6.45% accuracy before leveling off. By contrast, Exacta appears to be a poor bet, with AZGreyhound losing money for cutoffs above 1.6. Both Superfecta Box and Trifecta Box garnered substantially higher payouts than the other exotic wagers for two reasons. First, since the odds of correctly selecting the Trifecta and Superfecta combinations are markedly low, these types of wagers will inherently have higher payouts. Second, in spite of these low odds of correct selection, AZGreyhound is able to choose the correct winning set in a consistent manner. This tradeoff between accuracy and payout can further be illustrated in betting efficiency as shown in Figure 9.

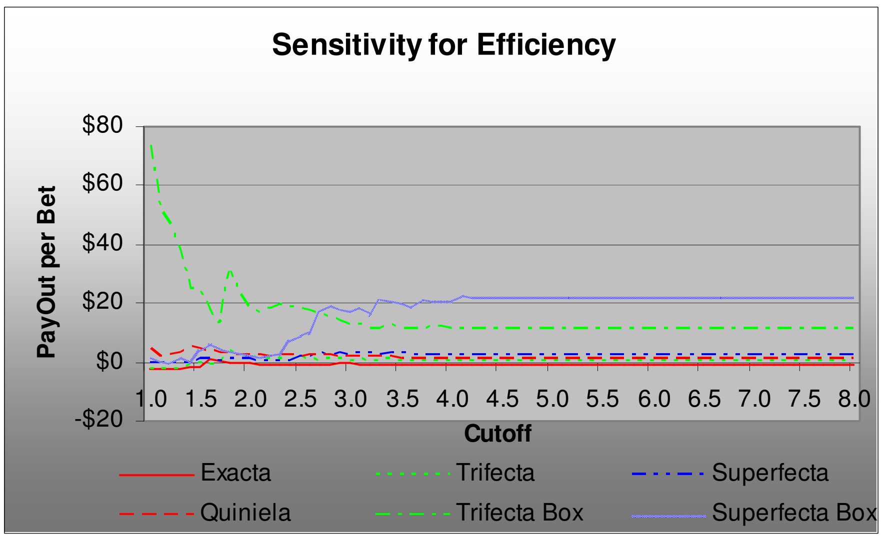

Figure 9. System Efficiency for Exotic Wagers

From this figure, both Trifecta Box and Superfecta Box had the highest efficiencies. Trifecta Box peaked at a $73.57 return per $2 bet at cutoff 1.0 and Superfecta Box peaked at $21.96 for cutoff 4.1 and leveled out to $21.20 for cutoffs 5.0 and higher. The positive return would imply that the system was profitable with longshot combinations.

Furthermore, we decided to dive deeper into the data and explore the reasons behind why AZGreyhound is able to make consistently better predictions. Table 7 examines Superfecta Box wagering broken down by the various tracks.

Track	Accuracy	Payout	Efficiency	Std Dev
Caliente	7.98%	**$17,353.20**	$37.40	$322.88
Lincoln	8.89%	$11,850.80	**$64.41**	$379.16
Raynham-Taunton	6.22%	$10,126.23	$33.31	$177.07
Tucson	10.53%	$1,534.16	$7.04	$78.44
Wichita	**14.62%**	$1,423.67	$9.07	$58.91

Table 7. Superfecta Box Wagering by Greyhound Track

In Table 7, AZGreyhound had the highest accuracy predicting Superfecta Box wagers in Wichita, 14.62% as compared to random chance at 2.79% accuracy. This would suggest that there exists a larger gulf between *winners* and *losers* at Wichita than at other tracks. Caliente has the highest payout at $17,353.20. This comes from Caliente running more races than other tracks, however, Caliente has a high standard deviation which implies short bursts of high paying wagers. Wichita had the lowest standard deviation meaning that payout returns, while on average low at $9.07, are more uniform in distribution. Lincoln has the best Efficiency per bet at $64.41 return. While Lincoln has fewer bets than other tracks, those bets are netting larger longshot pots. Breaking the data down by the day of the week also nets some interesting results as shown in Table 8.

Day of Week	Accuracy	Payout	Avg	Std Dev
Sunday	6.19%	$6,513.51	$24.40	$317.17
Monday	7.91%	$4,127.58	$17.49	$146.12
Tuesday	5.70%	$4,379.14	$25.76	$288.29
Wednesday	**12.28%**	$7,327.74	**$28.08**	$221.02
Thursday	12.14%	$1,832.13	$8.00	**$67.78**
Friday	12.05%	$5,580.39	$15.63	$130.33
Saturday	11.37%	**$11,637.89**	$26.94	$221.68

Table 8. Superfecta Box Wagering by Day of the Week

Table 8 shows that Wednesday has the highest AZGreyhound accuracy of 12.28% as well as the highest Efficiency of $28.08 payout return per bet. We believe that because a good proportion of the tracks do not race on Sunday through Tuesday, that greyhounds have time to rest up before a Wednesday race and hence are more predictable. Thursday has the most uniform distribution of payouts with a standard deviation of $67.87. Saturday has the highest payout, $11,637.89, however, it also has the most races of any day. If we were to further break down the data by track and day of the week, we would have the results shown in Table 9.

Track	Day of Week	Accuracy	Payout	Efficiency
Caliente	Sunday	6.19%	**$5,818.60**	$63.94
Caliente	Monday	8.62%	$2,236.20	$29.82
Caliente	Tuesday	7.69%	$4,489.20	$54.75
Caliente	Wednesday	8.55%	$718.00	$16.70
Caliente	Thursday	8.40%	$513.40	$12.22
Caliente	Friday	8.33%	$2,014.20	$50.35
Caliente	Saturday	8.33%	$1,563.60	$17.18
Lincoln	Friday	9.17%	$2,382.40	$34.53
Lincoln	Saturday	8.89%	$5,728.60	**$84.24**
Raynham-Taunton	Friday	7.09%	$1,399.89	$15.22
Raynham-Taunton	Saturday	6.15%	$3,393.54	$46.49
Tucson	Monday	7.07%	$191.21	$6.17
Tucson	Tuesday	3.60%	$5.74	$0.15
Tucson	Wednesday	10.83%	$1,129.50	$31.37
Tucson	Thursday	15.45%	$326.04	$10.52
Tucson	Friday	9.23%	-$27.13	-$0.62
Tucson	Saturday	12.50%	-$91.20	-$2.40
Wichita	Wednesday	14.15%	-$60.00	-$2.40
Wichita	Thursday	12.24%	$276.40	$11.52
Wichita	Friday	**17.86%**	$4.63	$0.17
Wichita	Saturday	15.49%	$1,054.20	$22.43

Table 9. Superfecta Box Wagering by Track and Day of the Week

From this table, Wichita on Fridays has the highest accuracy of 17.86%. Caliente on Sundays has the highest payout of $5,818.60 and Lincoln on Saturdays has the highest betting efficiency of $84.24. However, a closer look at Caliente on Sundays shows that Feb 18 was an abnormal day, as shown in Table 10.

Date	Payout
2/4/2007	-$28.80
2/11/2007	-$45.60
2/18/2007	$5,138.80
2/25/2007	$451.40
3/4/2007	$302.80

Table 10. Superfecta Box Payout for Caliente on Sundays

7. Conclusions and Future Directions

Within traditional wagers the Show bet appeared the best. Show had higher accuracy followed by Place and Win or all cutoffs. AZGreyhound's picks for Win, Place and Show were all significantly better than random chance. Show also demonstrated higher payouts and betting efficiency than Place and Win for cutoffs above 1.8. This stems from AZGreyhound picking greyhounds with longer odds and subsequently the higher payouts.

For straight wagering, Exacta, Trifecta and Superfecta, AZGreyhound's picks were all significantly better than random chance. Exacta had the highest accuracy for cutoffs above 1.5 and Superfecta had higher payout and efficiency returns for cutoffs above 2.4. This is also the result of AZGreyhound able to capitalize on the longer odds more accurately than random chance alone.

For box wagering, Quiniela had the highest accuracy for all cutoffs above 1.3. AZGreyhound's picks for Quiniela, Trifecta Box and Superfecta Box were all significantly better than random chance. Superfecta Box had the highest payout and efficiency for cutoffs above 2.7. Again this is the result of AZGreyhound able to capitalize on the longer odds more accurately than random chance alone. When betting Superfecta Box on every race, regardless of cutoff, accuracy was 6.35%, well above random chance at 2.79%.

While this system demonstrates a marked promise of better prediction, the reader should be cautioned that the act of making large bets on races will change the race odds to the detriment of the bettor. Similarly, like the Dr. Z system, should a significant enough population begin to engage in SVR prediction, any gains will be effectively arbitraged away.

Further research could include adopting the SVR algorithm to the problem of similar sport-related predictions including thoroughbred and harness racing as well as more mainstream sports such as baseball.

References

Arrow, K., (1965), Aspects of he Theory of Risk Bearing. Helsinki, Finland, Yrjo Jahnsson Foundation.

Burns, E., R. Enns and D. Garrick, (2006), The Effect of Simulated Censored Data on Estimates of Heritability of Longevity in the Thoroughbred Racing Industry. *Genetic Molecular Research* 5(1), 7-15.

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

Chen, H., P. Rinde, L. She, S. Sutjahjo, C. Sommer and D. Neely, (1994), Expert Prediction, Symbolic Learning, and Neural Networks: An Experiment on Greyhound Racing. *IEEE Expert* 9(6), 21-27.

GRAA, (2008), Greyhound Racing Association of America. Retrieved on April 29, 2008, from http://www.gra-america.org

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

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

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

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

Philpott, A., S. Henderson and D. Teirney, (2004), A Simulation Model for Predicting Yacht Match Race Outcomes. *Operations Research* 52(1), 1-16.

Platt, J. C., (1999), Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods: Support Vector Learning. B. Scholkopf, C. Burges and A. Smola. Cambridge, MA, MIT Press: 185-208.

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

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

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

Schumaker, R. P. and H. Chen, (2006). Textual Analysis of Stock Market Prediction Using Financial News Articles. *Americas Conference on Information Systems*, August, Acapulco, Mexico.

Schumaker, R. P. and H. Chen, (2008), Evaluating a News-Aware Quantitative Trader: The Effects of Momentum and Contrarian Stock Selection Strategies. *Journal of the American Society for Information Science* 59(1), 1-9.

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.

Tay, F. and L. Cao, (2001), Application of Support Vector Machines in Financial Time Series Forecasting. *Omega* 29, 309-317.

Vapnik, V., (1995), The Nature of Statistical Learning Theory. New York, Springer.

Witten, I. H. and F. Eibe, (2005), Data Mining: Practical Machine Learning Tools and Techniques. San Francisco, Morgan Kaufmann.

Ziemba, W. and D. Hausch, (1984), Beat the Racetrack. San Diego, Harcourt, Brace & Jovanovich.