Horse Racing Prediction at the Champ De Mars using a Weighted Probabilistic Approach

Sameerchand Pudaruth
University of Mauritius

Nicolas Medard
University of Mauritius

Zaynah Bibi Dookhun
University of Mauritius

ABSTRACT

Horse racing is a popular sport in Mauritius which attracts huge crowds to Champ de Mars. Nevertheless, bettors face many difficulties in predicting winning horses to make profit. The principal factors affecting a race were determined. Each factor, namely jockeys, new horses, favourite horses, previous performance, draw, type of horses, weight, rating and stable have been examined and appropriate weights have been assigned to each of them depending on their importance. Furthermore, data for the whole racing season of 2010 was considered. The results of 240 races of 2010 have been used to determine the degree to which each factor affect the chance of each horse. The weights were then summed up to predict winners. The system can predict winners with an accuracy of 58% which is 4.7 out of 8 winners on average. The software outperformed the predictions made by the best professional tipsters in Mauritius who could forecast only 3.6 winners out of 8 races.

Keywords
Betting, Champs de Mars, horse racing, prediction, statistics.

1. INTRODUCTION
Betting and winning is a skill which needs to be developed with proper analysis of all the factors that can affect a particular race. It cannot be denied that horseracing has been inculcated in the Mauritian culture since the time that it was introduced in Mauritius. The influence of media has also contributed to its significant growth. Over the years, betting has become more complex and widespread [1]. Therefore, a great deal of effort is needed to understand the major principles of betting. As a matter of fact, Mauritian bettors often face difficulties in forecasting the winner of each race and thus end up making a loss.

Large amounts of data are published about races every week to which the general public has easy access. In addition, professional tipsters in Mauritius give out tips about possible winners of each race through newspapers, sports channels and the various racing websites. Even though the professional tipsters may probably have access to more information about races than the public, their tips are often wrong.

There are many factors that affect a race and previous data have to be analysed over a period of time to find any trends. Common people find it very difficult to gather loads of data for analysis so as to predict a winner with high accuracy. Thus, our work is geared towards finding the maximum number of winners per week. This can be possible by designing a computer program to collect and evaluate the large amount of data available about horse racing and the use it to predict the results of races.

The aim is to maximise the profit made by gamblers even if they have limited knowledge about horse racing. It is important that the solution to be devised to be more reliable than the tips provided by professional tipsters. Using a computer program, the factors influencing the outcome of a race will be thoroughly analysed to make good predictions. Data for the 2010 racing season has been taken and populated in a database. This will help to find trends among the data and this will be used to predict winners in the future.

Numerous studies have been carried out by researchers to find out a way to predict the results of different sports. This will be further discussed in the next section.

2. LITERATRE REVIEW
Prior to betting, people analyse the horses and jockeys to find out the possible winners and this is mainly done by using the tips on winning horses which are often available on certain websites, newspapers and television sports channel. But these tips are far from providing accurate predictions to the numerous bettors across Mauritius. Many people around the world have been studying the horse racing business and have subsequently developed several algorithms to forecast winning horses accurately. In this review, the approaches used by different researchers to make their predictions on different sports have been investigated and discussed.

Williams and Li studied horse racing in Jamaica in order to find probable winners [2]. The main methodology used was Artificial Neural Network, an adaptive system that is trained from existing data and can predict results when offered with new input patterns. In Jamaica, thoroughbred horses race over distances of 1000m to 3000m. Data from 143 races were collected from 1st January to 16th June 2007. Four different learning algorithms were applied to the data and the results obtained were compared. Each algorithm had to find patterns in the data provided and forecast the winners in each race. They studied different variables connected to horse racing such as racing distance, past position, weight of horse and jockey, horse's finishing time and others as input to the Neural Network. 80% of the data collected was used to train the system and the remaining 20% was used for testing purposes. Their investigations showed that all the four algorithms predicted correct horses with an accuracy of about 74% and they produced more or less similar results.

Baulch investigated the use of Artificial Neural Network in rugby league and basketball games [3]. He used two types of neural networks to predict winners in both games. The data

used for analysis were the team's previous performances, the number of wins and the average scores per game to predict the winner. He found that the results of the trained networks produced an accuracy that lies between 55% and 58.2% for rugby and between 49% and 59% for basketball. He concluded that neural networks helped a great deal in sports predictions and better results can be obtained from more in-depth research.

Chen et al. studied greyhound racing. Initially, they found it very difficult to apply neural network algorithms as there are about 50 different variables available to bettors on race cards [4]. The neural network cannot use all these variables to predict accurate results and thus they decided to filter them out and analyse only what they thought to be the most important variables. Data from 200 races and 1600 greyhounds were gathered for training. An additional 100 races with 800 greyhounds were used for testing purposes. The system forecasted winners for 34 races, but wrongly predicted 50 results and no winners were predicted for 16 races. They placed bets on dogs that were predicted winners by the system. The final payment for 100 races were $69.20 whereas when three human experts have predicted winners the final payoff were -$71.40, -$61.20 and -$70.20, respectively. These values clearly showed that the neural network predictions outperformed those of the human experts.

Davoodi and Khanteymoori made horse racing predictions for a race track in New York using neural networks [5]. Data for 100 races were collected for analysis from 1st to 29th January 2010. Artificial Neural Networks were used to forecast the finishing time of each horse in a race. Supervised learning has been used where the neural network was provided with training data and the desired output. Five supervised algorithms were used in their study to compare their forecasting power and performances. Eight variables had been used as input: Horse weight, type of race, horse's trainer, jockey, number of horses in a race, race distance, the track condition and weather. These data were input and one neural network was used for each horse. The finishing times along with the horses' ranks were then predicted. All the algorithms used had an average accuracy of about 77%.

Pardee investigated the use of neural networks to predict outcomes of football matches in college [6]. He developed a model using data from past football seasons to find out the winning team. He examined data from 11 games for each team which played in the 1998 and 1999 seasons. He used those historical data to train and test the system. The system developed could predict which team will win the college football game with an accuracy of 76%. Upon comparing the results of his predictions with those of a television channel, he found that the latter's forecast were only accurate up to 68.2%. He also stated that if given more data about the team's previous performances, the degree of accuracy of his system could have been higher.

McCullagh examined the use of neural networks to support the data mining process for the talent identification problem to select potential players for the annual Australian Football League [7]. He used a data set of 386 players and 58 attributes which have been tested from 1999 to 2005. He carried out an analysis to compare the results from recruiting managers and the neural network. Data for 20 players and recruiting managers from 16 different clubs were selected for the study. However, it was found that in most of the cases, the recruiting managers gave the most correct predictions (66.8% accuracy)

compared to an accuracy of 60.1% from the neural networks. This can be explained by the fact that the neural network depend data like body composition, flexibility, anaerobic and aerobic power, visual tests, skill assessments and subjective assessments on strengths, weaknesses and personal attributes of the players while the recruiting managers have widespread networks of people watching games of prospective players over many years, interviews and videos as well. Hence, he could only conclude that the neural network can be used by the recruiting managers to support and enhance their decisions before choosing players but cannot replace the expertise of the recruiting managers.

3. DESIGN AND PROPOSAL

Data for the racing season of 2010 has been collected and populated in a database. Several interviews have been carried out with people from the horse racing business to find out the factors that affect a race. The results of 240 races in 2010 have been thoroughly examined to find out to which extent the factors affect a race. Furthermore, all the data have been scaled to a range between 0 and 1 and weights have been assigned to each factor.

A software was developed to enable users to input details about new races and to make predictions about potential winners. The factors that affect a race are discussed below as well as the algorithm that has been used to allocate weights to them.

Factor 1: Jockey

The jockey is an important determinant that greatly affects the winning chance of a horse. The performance of each jockey in a race varies. The weights attributed to each jockey will be the total wins of the jockey divided by its total mount.

$$Factor\ 1 = \frac{total\ wins}{total\ mount}$$

Factor 2: New horse

An interesting trend has been found when analysing past results of races over 3 consecutive years. On average, only 12% of new horses win on its first race in Mauritius. Thus, it can be assumed that a new horse has 12% chance of winning the race and the others have 88%.

$$Factor\ 2, new\ horse = 0.12$$

$$Factor\ 2, not\ a\ new\ horse = 0.88$$

Factor 3: Odds

The odds at which a horse is offered can be very indicative of its actual chance of winning. Normally, bookmakers will offer the lowest return on a horse which according to them has the highest chance of winning the race. Therefore, as a general principle, the lower the odds of a horse, the higher is its chance of winning.

Furthermore, it has been concluded from the research made on the results of 3 consecutive years that on average 33.06% of favourite horses win a race. Thus, the algorithm to be applied is: find the lowest odd and divide the odds of each horse by the lowest odd and multiply by 0.3306.

$$Factor\ 3 = \left(\frac{lowest\ odd}{odds\ of\ horse}\right) * 0.3306$$

Factor 4: Previous performances

This factor takes into consideration the previous performances (i.e. rank) of all the horses taking part in a race and tries to assign numerical values to this factor. We consider only the last three performances of each horse. The algorithm is as follows:

$$Factor\ 4 = Rank3 * 1 + Rank2 * 2 + Rank1 * 3$$

Rank1 is the rank of the horse in the last race that it ran, etc. The algorithms proceeds as follows: (i) compute a total for each horse in a race (ii) find the lowest total (iii) divide the lowest total by the total of each horse.

Factor 5: Draw

The draw is the starting stall on which a horse has been placed and it is a determining factor. The idea here is to study the frequency of winning of horses starting from particular draw in relationship to the number of horses participating in the race.

For example, calculating the number of times that a horse starting from draw 1, wins the race when there are 8 horses. The percentage must be found out and assigned to the horses when they take part in a race, depending on their draw and number of horses taking part in the race.

Factor 6: Type of horse

Some of the horses run better on some particular distances and do not perform well on other distances. The distances are measured in meters. Thus, the horses are classified in five distance categories as shown below:

Category 1: 1000

Category 2: 1365-1400

Category 3: 1500-1600

Category 4: 1850-2050

Category 5: >2050

However, it is possible that a horse is classified in more than one category. The classification of each horse is done based on its previous performances on different distances. A horse will be classified in a category in which it performs best. It has been noted that on average, a horse running on distance of its category wins **87%** of the time. Thus, the algorithm is shown below:

$$Horse\ running\ in\ its\ category = 0.87$$

$$Horse\ not\ running\ in\ its\ category = 0.13$$

Factor 7: Weight

The lowest weight that can be carried by a horse during a race is 49 kg while the highest weight is 62 kg. It is reasonable to assume that as the weight gets down, a horse will generally perform better.

$$Factor\ 7 = \frac{lowest\ weight}{weight\ of\ horse}$$

Factor 8: Rating

The rating is an already computed value which shows the strength of a particular horse. Thus, by principle, the higher the rating of a horse is, the stronger it is. Thus, the algorithm is to divide the rating of each horse by the highest rating among the horses participating in the race.

Factor 9: Stable

Some stables have better reputation than others because their horses win more races than the others. This is usually due to better treatment and better training. Thus, the algorithm is to take the percentage wins of each stable and assign it to the horse.

Worth of Each Horse

The value that will allow a horse to be predicted as winner is computed by summing up the values return by each factor.

Total = Factor 1 + Factor 2 + Factor 3 + Factor 4 + Factor 5 + Factor 6 + Factor 7 + Factor 8 + Factor 9

For each horse participating in the race, there will be a total. Thus, the horse with the highest total is the predicted winner.

4. IMPLEMENTAION

The software has been designed on Java which will enable users to enter data about horses, previous races, stables and jockeys through a GUI. These details will be stored into a database and the user can then use them to predict the winner of each race. The user can also view previous results of races.

When the option "Predict Result" is used, the system will fetch the corresponding weights from the database. The winner of each race is determined by calculating the total of the different weights. The winner is the horse with the highest total. The results are stored in a database. The weights are re-adjusted automatically as soon as new data is entered.

The following example shown demonstrates how the system will calculate the value for each factor for race 313 (the 3rd race of the 31st meeting). We can see from Table 1 that the

Table 1. Weights assigned to each horse of Race 313 and the Total

Horse	Jockey	New horse	Odds	Draw	Category	Previous Performance	Weight	Rating	Stable	Total
CAPTAINS CHEER	0.2549	0.88	0.3306	0.11111	0.86667	1	0.81667	1.02222	0.19643	5.4786
UMHLANGA	0.33333	0.88	0.13987	0.10794	0.86667	0.33333	0.85965	0.86667	0.08955	4.47701
BAYNESFIELD	0.15909	0.88	0.09092	0.08889	0.86667	0.2	0.83051	0.84444	0.09005	4.05057
WAR ACADEMY	0.18182	0.88	0.18183	0.11429	0.86667	0.25	0.83051	0.84444	0.14607	4.29563
BALLITO BOY	0.1	0.88	0.24244	0.10476	0.86667	0.33333	0.82353	0.84444	0.07516	4.27033
LIZARDS LUST	0.0625	0.88	0.01515	0.11111	0.86667	0.11111	0.86726	0.73333	0.07516	3.72229
RUM RATION	0.06	0.88	0.02273	0.06032	0.13333	0.16667	0.85217	1.04444	0.07333	3.29299
TANGERINE	0.20946	0.88	0.03953	0.08254	0.86667	0.16667	0.88288	0.62222	0.19048	3.94045
JUST OPPOSITE	0.03774	0.88	0.0202	0.09524	0.86667	0.125	0.89091	0.66667	0.08955	3.67198
POWERFUL GIFT	0.10169	0.88	0.01818	0.09206	0.13333	0.11111	0.83051	0.8	0.13924	3.10612

horse Captains Cheer has the highest total. Thus, our system will predict Captains Cheer as the winner.

The values in Table 1 have been sorted in descending order of Total to ease understanding. At the same time, it is also possible to get an idea of the likely rank of the other horses. A graph of the horse against the prediction values in the Table 1 is shown below.
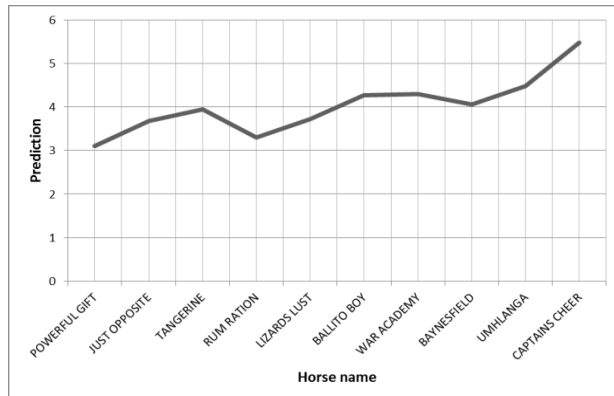

Fig. 1 Graph of horse against prediction values of race 313

5. EVALUATION

The algorithm described above has been applied on 24 races (3 consecutive horse racing meetings). The result obtained after using the software was 58.33% success. It should be noted that 100 % prediction is not possible in horse racing due to the existence of unexpected factors that cannot be predicted for, e.g. a last minute change in weather, a horse may be sick on the day of the race or a jockey may fall off the horse while racing. On average, the software has predicted 4.7 winners out of 8 on a racing day whereas the best professional tipsters could only forecast 3.6 out of 8 winners.

The results obtained by the system clearly outperform the prediction made by the best tipsters in Mauritius who predicted winners with a success of only 44% for the same period of time. The calculations and predictions by the system have been done using data available publicly and it has achieved accurate results compared to those tips even though they might have more information at their disposal and possibly inside information as well. The system is designed to be used on computers where calculations are done within seconds.

An analysis has been performed to see how much profit a player can make. Assuming that the player bets Rs100 on each of the 24 races considered above, it was found that his average profit per race was Rs64. Thus, if he increases his bet to Rs1000 per race, he can expect to get a profit of Rs640 per race. This is a very significant amount as this amount to Rs5120 per racing day. Thus, our system enables a person to make a profit by betting on horses, without even having to spend precious time following training of horses at the Champs de Mars and without the hassle of keeping track of so much information.

6. CONCLUSION

The Champs de Mars horse racing track is unique in the world, in the sense that it is almost circular and the final straight is very short, only a mere 300m compared to a standard of at least 600m worldwide. Furthermore, the track is not flat, i.e., it has a certain inclination. New horses, new jockeys and even new stables are common things at the Champs de Mars. Thus, it is quite difficult to predict the outcome of races in Mauritius. Professional tipsters could predict only 3.6 winners on average out of 8 races for the whole racing season of 2010. Our horse racing prediction software forecasted 4.7 winners out of 8 races. However, our testing was performed only on three meetings. The strength of our tool is that as new data about races are input into the system, all the weights are automatically readjusted for more accurate predictions in the future. Therefore, we see that our prediction system is more reliable than professional tipsters. Nevertheless, we believe that it is still possible to further increase the percentage success if additional factors like age and equipment were considered. For the future, we intend to apply neural networks, decision trees and fuzzy logic on our data to see how it may improve our predictions.

7. REFERENCES

[1] Mauritius Turf Club [online]. (2012) [Accessed 21 April 2012].

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

[3] Baulch, M. "*Using Machine Learning to Predict the Results of Sporting Matches*". Department of Computer Science and Electrical Engineering, University of Queensland, 2001.

[4] Hsinchun Chen, Peter Buntin Rinde, Linlin She, Siunie Sutjahjo, Chris Sommer, Daryl Neely, "*Expert Prediction, Symbolic Learning, and Neural Networks: An Experiment on Greyhound Racing*," IEEE Intelligent Systems, vol. 9, no. 6, pp. 21-27, Dec. 1994.

[5] Davoodi E., Khanteymoori,m A. R. "*Horse Racing Prediction Using Artificial Neural Networks*". NN'10/EC'10/FS'10 Proceedings of the 11th WSEAS International Conference on Neural Networks.

[6] Pardee, M., "*An Artificial Neural Network Approach to College Football Prediction and Ranking*". Technical paper. Madison, WI: University of Wisconsin, Electrical and Computer Engineering Department, 1999.

[7] McCullagh, J. "*Data Mining in Sports: A Neural Network Approach*". International Journal of Sciences and Engineering, (3): 131-138, 2010.