Be part of a better internet. Get 20% off membership for a limited time

✦ Member-only story

Bill Benter (1994) Computer Based Horse Race Handicapping and Wagering Systems: A Report

A summarised article to let you have a basic understanding of parimutuel handicapping model

Yeung WONG · Follow

Published in Parimutuel Racetrack Analysis · 8 min read · Dec 28, 2019
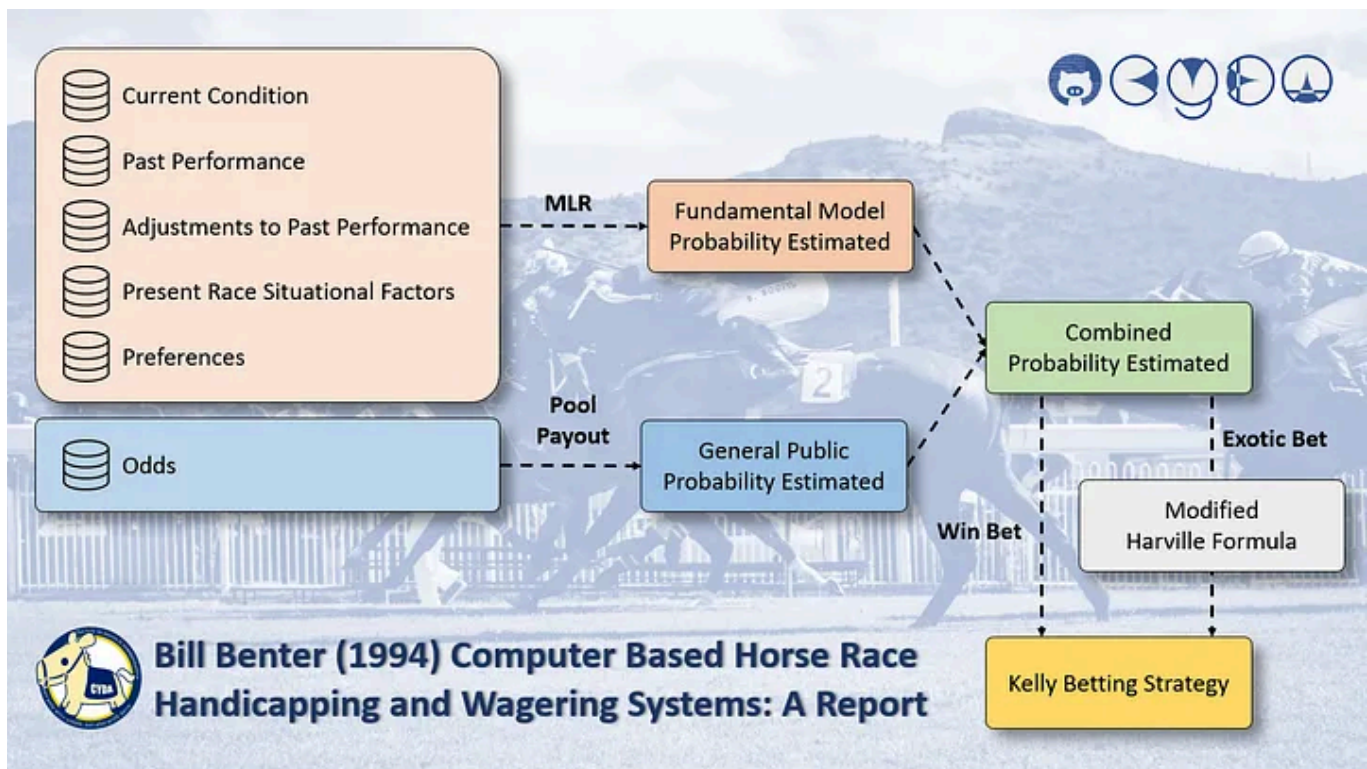
👏 107   💬 1

Photo by cyda

Summary

This paper examines the elements necessary for a practical and successful computerized horse race handicapping and wagering system. Data requirements, handicapping model development, wagering strategy, and feasibility are addressed. A logit-based technique and a corresponding heuristic measure of improvement are described for combining a fundamental handicapping model with the public's implied probability estimates. The author reports significant positive results in five years of the actual implementation of such a system. This result can be interpreted as evidence of inefficiency in pari-mutuel racetrack wagering. This paper aims to emphasize those aspects of computer handicapping which the author has found most important in the practical application of such a system.

Motivation

Certain authors are concerned mainly with the public odds and do not attempt to predict horse performance from fundamental factors.

1. *Dr. Z's Beat the Racetrack (Ziemba, William T. and Donald B. Hausch, 1987)*

2. *Betting at the Racetrack (Ziemba, William T. and Donald B. Hausch, 1986)*

3. *Racetrack Betting: The Professors' Guide to Strategies, (Asch, P., RE. Quandt, 1986)*

Although this requires relatively little preparatory effort and can be effectively employed by the occasional racegoer, the betting opportunities tend to occur infrequently and the maximum expected profit achievable is usually relatively modest. **Therefore, Bill Benter decided to develop a computer model to estimate each horse's probability of winning and calculate the appropriate amount to wager.**

Advantages of Computer Approach Handicapping

1. Due to **empirical nature**, one need not possess specific handicapping expertise to undertake this enterprise, as everything one needs to know can be learned from the data.

2. The **testability** of a computer system is unquestionable. One can develop a model and test it on unseen races and obtain a reasonable estimate of the system's real-time performance before wagering any actual money.

3. The **consistency** of a computer system regardless of the mental state of the operator. This is a non-trivial advantage considering that a

professional level betting operation may want to bet several races a day for extended periods.

Features Engineering

Bill Benter classified various types of factors into five groups.

1. **Current Condition**
   - Performance in recent races
   - Time since last race
   - Recent workout data
   - Age of horse

2. **Past Performance**
   - Finishing position in past races
   - Lengths behind the winner in past races
   - Normalized times of past races

3. **Adjustments to Past Performance**
   - Strength of competition in past races
   - Weight carried in past races
   - Jockey's contribution to past performances
   - Compensation for bad luck in past races
   - Compensation for advantageous or disadvantageous post position in past races

4. **Present Race Situational Factors**
   - Weight to be carried
   - Today's jockey's ability
   - Advantages or disadvantages of the assigned post position

5. **Preferences which could Influence the Horse's Performance in Today's Race**

- Distance preference

- Surface preference (Turf vs Dirt)

- Condition of surface preference (Wet vs Dry)

- Specific track preference

Race Data Required

Bill Benter suggested the **minimum amount of data needed** for adequate model development and testing samples is in the range of **500 to 1000 races** since out-of-sample predictive accuracy does not seem to improve dramatically with development samples greater than 1000 races.

Features Selection

It is important to **define factors that extract as much information** as possible out of the data in each of the relevant areas. By considering the following three different specifications of a 'distance preference' factor:

1. **NEWDIST (**Bolton and Chapman, 1986)

   This variable equals one if a horse has run three of its four previous races at a distance less than a mile, zero otherwise.

2. **DOK (**Brecher, 1980)

   This variable equals one if the horse finished in the upper 50th percentile or

within 6.25 lengths of the winner in a prior race within 1/16 of a mile of today's distance, or zero otherwise.

3. **DP6A** (Bill Benter's Model)

   For each of a horse's past races, a predicted finishing position is calculated via multiple regression based on all factors except those relating to distance. This predicted finishing position in each race is then subtracted from the horse's actual finishing position. The final magnitude of the estimate is standardized by dividing it by its standard error.

> *The best is the one that produces the greatest increase in predictive accuracy when included in the model.*
>
> *The general thrust of model development is to continually experiment with refinements of the various factors.*

Statistical Testing on Variables

Ideally, **each independent variable would capture a unique aspect of the influences affecting horse performance.** The trial and error method of adding independent variables to increase the model's goodness-of-fit results in the model tending to become a hodgepodge of highly correlated variables whose individual significances are difficult to determine and often counter-intuitive. For example, **"Number of Past Races"** is one of the more significant factors in the handicapping model and contributes greatly to the overall accuracy of the predictions although there is no common sense reason why this factor should be important.

Insiders' Betting Reflects in Odds

There will always be a significant amount of **inside information** in horse racing that cannot be readily included in a statistical model. **Trainer's and jockey's intentions, secret workouts, whether the horse ate its breakfast**, and the like, will be available to certain parties who will no doubt take advantage of it. Their betting will be reflected in the odds.

> *Therefore, it is needed to find a way of combining the judgments of two experts — the fundamental model and the public.*

For a race with entrants $(1, 2, \ldots, N)$, the win probability of horse i is given by

$$c_i = \frac{exp(\alpha \ln f_i + \beta \ln \pi_i)}{\sum_{j=1}^{N} exp(\alpha \ln f_j + \beta \ln \pi_j)}$$

where c_i is the combined probability estimate, f_i is the out-of-sample fundamental model probability estimate and π_i is the public's implied probability estimate

The parameters α and β can be estimated by maximizing the likelihood function of the given set of past races $(1, 2, \ldots, R)$ with respect to α and β.

$$\hat{\alpha}, \hat{\beta} = \arg\max_{\alpha, \beta} \prod_{j=1}^{R} c_{ji}$$

The estimated values of α and β can be interpreted roughly as the relative correctness of the model's and the public's estimates. The greater the value of α, the better the model.

Model Evaluation

R^2 is used to compare models and to assess the value of a particular model as a betting tool. Each set of probability estimates, either the public's or those of a model, achieve a certain R^2 defined as

$$R^2 = 1 - \frac{L(model)}{L(1/N_j)}$$

The R^2 value is a measure of the explanatory power of the model. An R^2 of 1 indicates perfect predictive ability while an R^2 of 0 means that the model is no better than random guessing.

Wagering Strategy

Kelly betting strategy balances the tradeoff between risk and returns. The Kelly strategy specifies the fraction of total wealth to wager so as to maximize the exponential rate of growth of wealth, in situations where the advantage and payoff odds are known. For a single bet,

$$K = \frac{Advantage}{Dividend - 1} = \frac{Expected\ Return - 1}{Dividend - 1} = \frac{Win\ Probability \times Dividend - 1}{Dividend - 1}$$

In Bill Benter's experience, betting the full amount recommended by the Kelly formula is unwise for the following reasons.

1. Accurate estimation of the advantage of the bets is critical. If one overestimates the advantage by more than a factor of two, Kelly betting will cause a negative rate of capital growth.

2. If it is known that regular withdrawals from the betting bankroll will be made for paying expenses or taking profits, then one's effective wealth is less than their actual current wealth.

3. Full Kelly betting is a 'rough ride', downswings during which more than 50% of total wealth is lost are a common occurrence.

> *Therefore, a fractional Kelly betting strategy is advisable, that is, a strategy wherein one bets some fraction of the recommended Kelly bet (e.g. 1/2 or 1/3).*

Advanced Wagering Strategy

However, **multiple bets** are available in each race, and each bet affects the final payoff odds, the exact solution requires maximizing a concave logarithmic function of several variables. For further discussion, please see

Growth Versus Security in Dynamic Investment Analysis (MacLean, LC., W.T. Ziemba and G. Blazenko, 1992)

Another constraint on betting is the effect that one's bet has on the advantage. In pari-mutuel betting markets, each bet decreases the dividend. Even if the bettor possesses infinite wealth, there is a maximum bet producing the greatest expected profit, any amount beyond lowers the

Open in app ↗

Medium   🔍 Search   ✎ Write   🔔   Ⓐ

Pari-mutuel Betting Models (Kallberg, J.G. and W.T. Ziemba, 1994)

Exotic Betting

Bill Benter indicated that the **Harville formula** used to specify the probability that the horses will finish in order was significantly biased and

should not be used for betting purposes as it will lead to serious errors in estimations.

Therefore, Bill Benter developed an effective and computationally economical way to correct for the bias.

Harville formula that specifies the probability that the horses will finish in order

$$\pi_{ijk} = \frac{\pi_i \pi_j \pi_k}{(1 - \pi_i)(1 - \pi_i - \pi_j)}$$

Bill Benter Modified Harville formula that specifies the probability that the horses will finish in order

$$\pi_{ijk} = \frac{\pi_i \sigma_j \tau_k}{(1 - \sigma_i)(1 - \tau_i - \tau_j)}$$

where the second array σ_i is

$$\sigma_i = \frac{exp(\gamma \ln \pi_i)}{\sum_{j=1}^{N} exp(\gamma \ln \pi_j)}$$

and the third array τ_i is

$$\tau_i = \frac{exp(\delta \ln \pi_i)}{\sum_{j=1}^{N} exp(\delta \ln \pi_j)}$$

The parameters γ and δ can be estimated by maximum likelihood estimation. Suggested by Bill Benter,

$$\hat{\gamma} = 0.81, \hat{\delta} = 0.65$$

> The parameter values used should **NOT** be considered to be universal constants as different racetrack would have significantly different values for the parameters.

Results

Bill Benter resulted in net profits for four out of five seasons and the loss incurred during the losing season was around 20% of starting capital.

1. Bill Benter used five man-years of effort to **organize the database** and **develop a handicapping model** that showed a significant advantage and another additional five man-years were necessary to **develop the operation** to a high level of profitability.

2. A sample of approximately **2000 races** was initially used for model development and testing.

3. A conservative **fractional Kelly betting strategy** was employed throughout, with wagers being placed on all positive expectation bets available in both normal and exotic pools except place and show bets.

4. Extremely **large pool sizes** (> USD $10,000,000 per race turnover) made for low volatility odds, therefore bets could be placed with **accurate estimations of the final public odds**.

5. Bets were made on all available races except for 5% races containing only unratable horses, resulting in **approximately 470 races** bet per year. The average track take was around **19%** during this period.

RESULTS
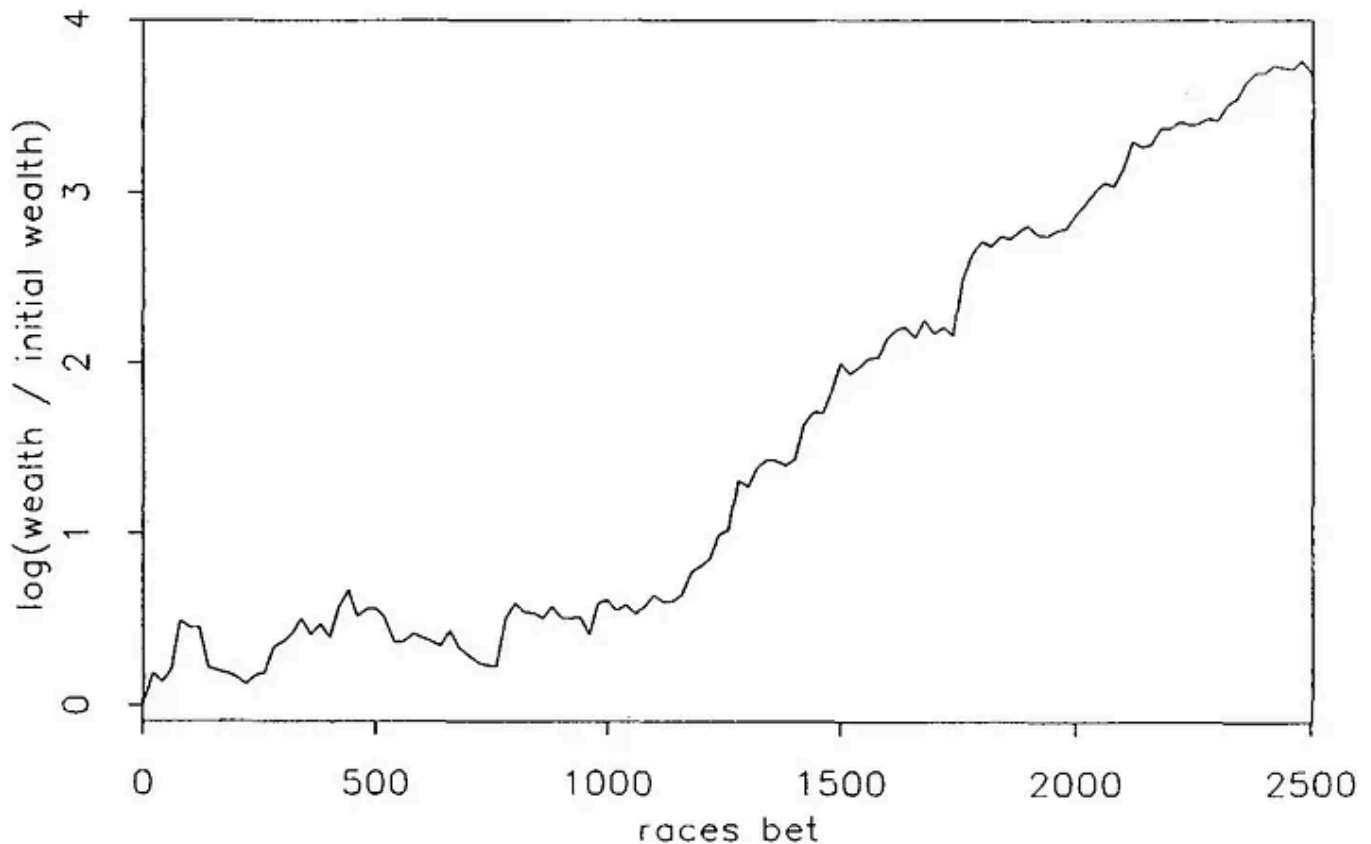
Words From Editor

This research paper is definitely the compulsory textbook for horse racing 101 course. For those beginners who wish to build a parimutuel prediction model, I **HIGHLY** recommend you to thoroughly read words after words.

There are no fancy or advanced formula that would hesitate you from kicking start the project. Better still, the project architecture are comprehensive and it is prefect for you to take as the reference guideline.

For me, the maximum gain would be understanding the idea of DP6A. I did not imagine using a residual of one model as another features in another

model. It really opens my eyes. This method is amazingly creative and practically very useful.

Another point interests me most will be how he leverages the public odds to build the second model. This perfectly solves the problem of Griffin horses. Before that, I only used the public odds to do the Kelly proportion. And now I learn how to adjust the estimated probabilities.

Data Science   Sports Betting   Horse Racing   Literature   Research

Written by Yeung WONG

Follow

898 Followers · Editor for Parimutuel Racetrack Analysis

Hong Kong | Master of Data Science | Hackathon Enthusiast | Public Speaker | Top Writter | R | Python | Sports Analytics | Investment | https://blog.cyda.hk/

More from Yeung WONG and Parimutuel Racetrack Analysis
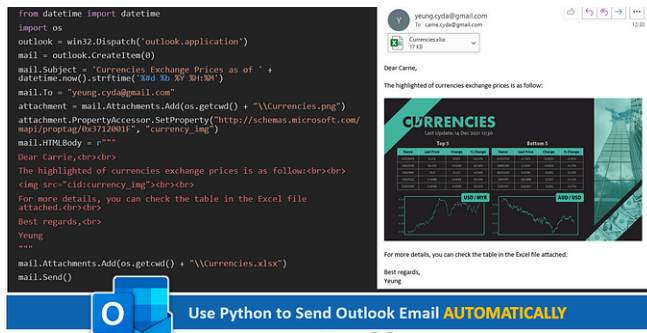
Yeung WONG

Use Python to Send Outlook Emails

Step by step tutorial to send notification email using Python

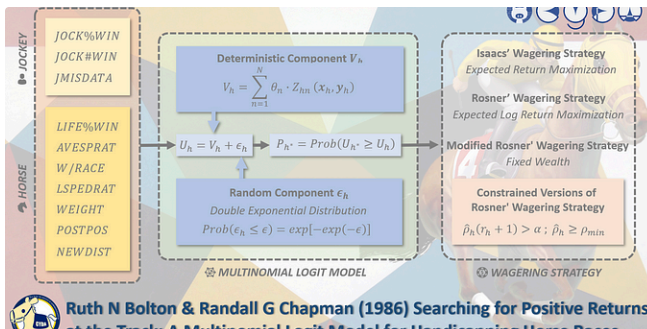⭐ Dec 16, 2021 · 👋 311 · 💬 2
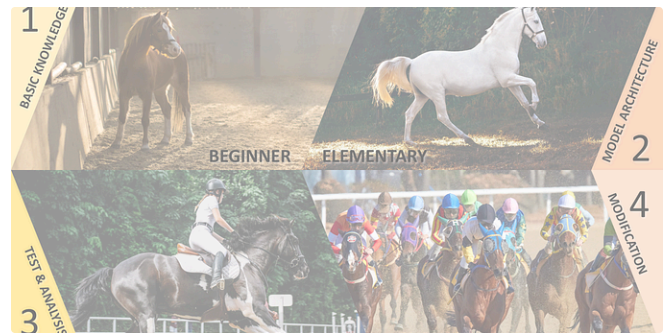
Yeung WONG in Parimutuel Racetrack Analysis

Horse Racing Research Papers

People always asked me "How can I start building a horse racing handicapping…
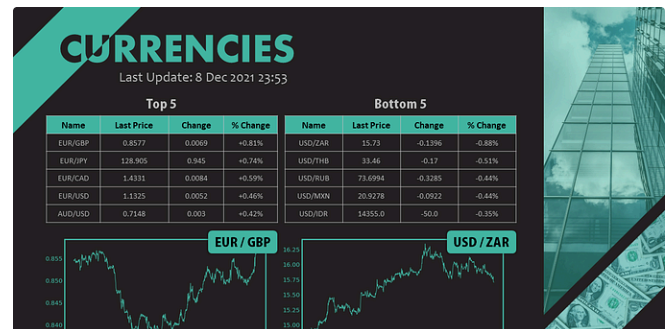
⭐ Jan 24, 2020 · 👋 61 · 💬 1

Yeung WONG in Parimutuel Racetrack Analysis

Ruth N Bolton & Randall G Chapman (1986) Searching for…

A summarised article to let you understand more about multinomial logit handicapping…

⭐ Jan 10, 2020 · 👋 101 · 💬 1

Yeung WONG in Towards Data Science

Use Python to Automate the PowerPoint Update

Step by step tutorial to edit PowerPoint slides using Python

⭐ Dec 9, 2021 · 👋 285 · 💬 5

See all from Yeung WONG

See all from Parimutuel Racetrack Analysis

Recommended from Medium

Haykaz Aramyan in LSEG Developer Community

Alexander Nguyen in Level Up Coding

How to build an end-to-end transaction cost analysis...

Overview

The resume that got a software engineer a $300,000 job at Google.

1-page. Well-formatted.

Mar 18

May 31   13.8K   206

Lists
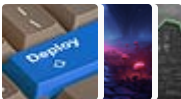
Predictive Modeling w/ Python

20 stories   ·   1388 saves

Practical Guides to Machine Learning

10 stories   ·   1678 saves

Coding & Development

11 stories   ·   703 saves

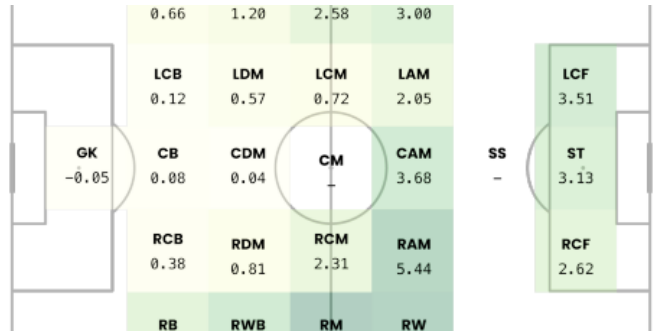### ChatGPT prompts

48 stories   ·   1804 saves
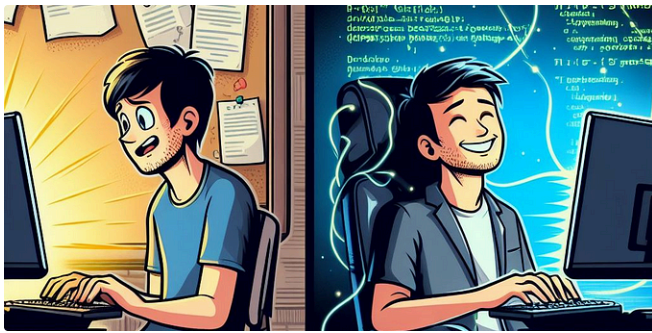
Abhay Parashar in The Pythoneers

17 Mindblowing Python Automation Scripts I Use Everyday

Scripts That Increased My Productivity and Performance

⭐ Jul 10   👏 3.4K   💬 28

Alex Marin Felices

Quantifying Creativity in Football: Introducing the Creative Decision…

A novel metric to quantify the creativity of soccer players' passes, based on both…

⭐ Mar 28   👏 10

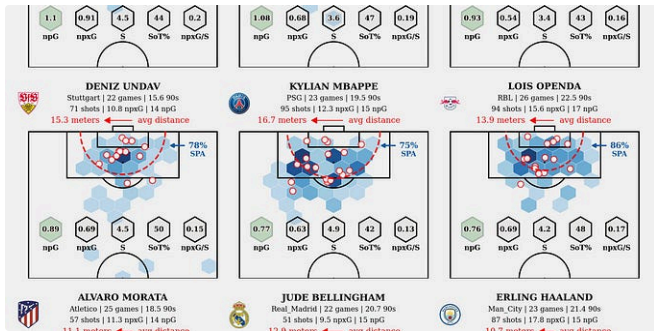Henrique Centieiro & Bee …   in Limitless Invest…

52% Returns in 30 Days: Your GPT-4o Quant Trading Bot Strategy

How to Profit in Any Market Condition with GPT-4o's Mean Regression Strategies

⭐ May 21   👏 1.93K   💬 32

Mikhail Borodastov

Plot Shots & Goals Maps with Python & MplSoccer. Ranking…

Step-by-Step Guide to Creating Comprehensive Shots and Goals Maps and…

Mar 18   👏 13

See more recommendations