



# TÓPICOS EM DATA SCIENCE PARA O ESPORTE

DIEGO RODRIGUES DSC

INFNET

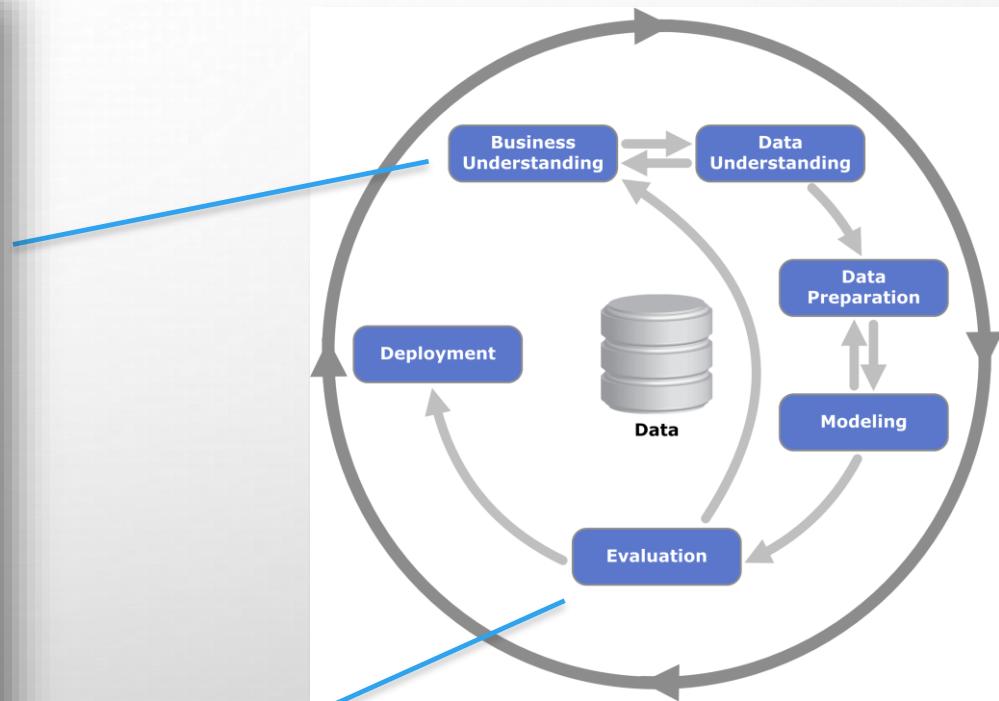
# AGENDA

- PARTE 1 : TEORIA
  - CRONOGRAMA
  - APRESENTAÇÃO DO PROFESSOR
- PARTE 2 : PRÁTICA
  - ORGANIZAÇÃO DO AMBIENTE
  - APRESENTAÇÃO DA TURMA

# **CRONOGRAMA**

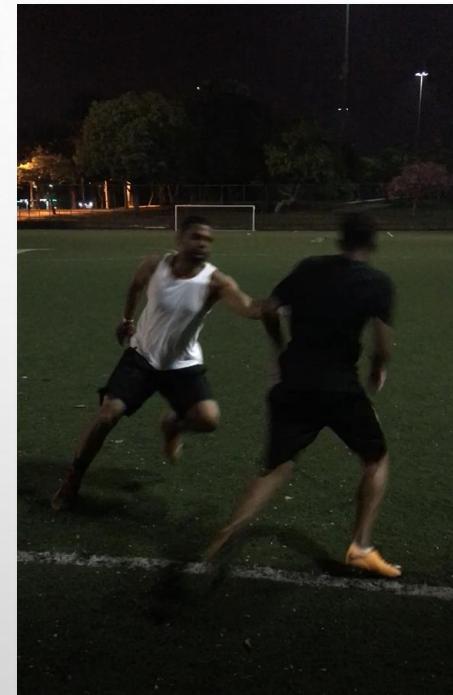
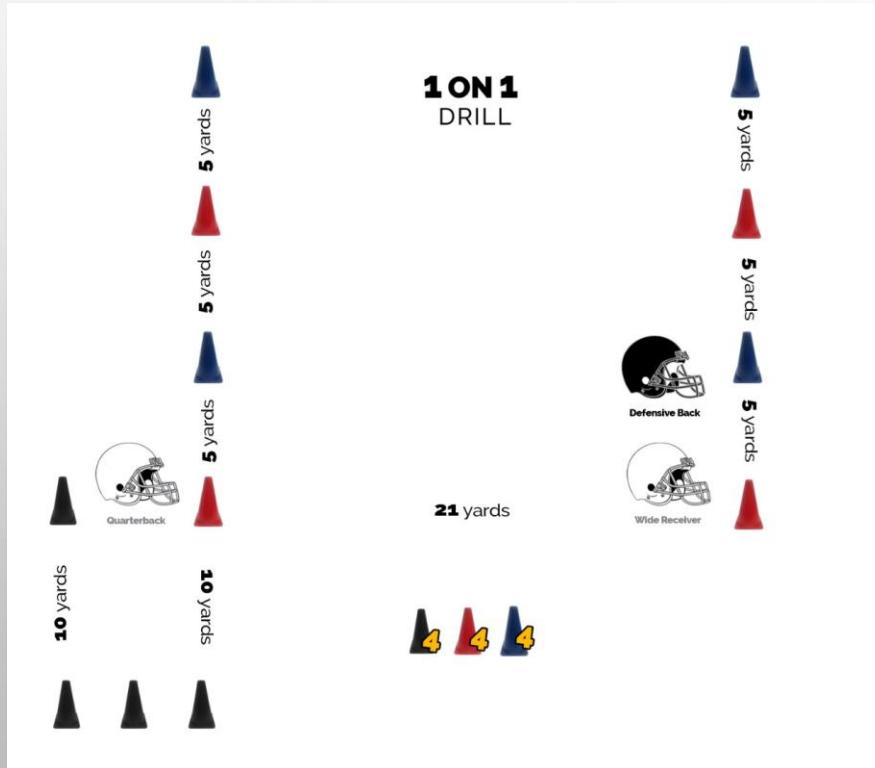
# CRONOGRAMA

DIA	NÚMERO	ÁREA	AULA	TRABALHOS
10/10/2023	1	Intro	Introdução a Disciplina e Organização do Ambiente	
17/10/2023	2	Dados	Coleta de Dados e Sensoriamento	
19/10/2023	3		Variáveis Aleatórias	Grupos
24/10/2023	4		Análise Exploratória	
26/10/2023	5		Estatísticas para Ranqueamento	
31/10/2023	6	Estatística	Rankeamento Estatístico : ELO	Base de Dados
07/11/2023	7		Rankeamento Estatístico : Glicko	
09/11/2023	8		Rankeamento Estatístico : TrueSkill	
14/11/2023	9		Rankeamento Estatístico : XELO	
16/11/2023	10		Modelos de Aprendizado de Máquina	Pesquisa
21/11/2023	11	ML	Machine Learning: Classificação	
23/11/2023	12		Machine Learning: Regressão	
28/11/2023	13		Machine Learning: Agrupamento	
30/11/2023	14		Machine Learning: Visão Computacional	Modelo
5/12/2023	15		Aplicações & Artigos: Esportes Independentes	
7/12/2023	16	Esportes	Aplicações & Artigos: Esportes de Combate	
12/12/2023	17		Aplicações & Artigos: Esportes de Objeto	
14/12/2023	18		Aplicações & Artigos : Betting	
19/12/2023	19	Workshop	Workshop	
21/12/2023	20		Apresentações de Trabalhos	Apresentação



CRISP-DM

# COLETA DE DADOS E SENSORIAMENTO



# COLETA DE DADOS E SENSORIAMENTO

GARMIN.

WEARABLES SPORTS & FITNESS OUTDOOR RECREATION AUTOMOTIVE MARINE AVIATION SALE

FREE GROUND SHIPPING ON ORDERS \$25 AND UP. FREE 2ND-DAY SHIPPING ON MOST ORDERS \$499 AND UP.\*

## MOST POPULAR SMARTWATCHES



Venu®



Forerunner®



Instinct®



EPIX™



fēnix®



vivoactive®

Style and design meet health and fitness.

For any reason to run.

Keep it bold, rugged and ready for adventure.

The metrics you need for epic performance.

Push the limits of multisport performance.

Stay healthy and active every day.

TODOS OS SENsoRES

SENsoRES DE FREQUÊNCIA CARDÍACA

SENsoRES DE MOVIMENTO



## Todos os Sensores

A excelência da tecnologia de monitoramento cardíaco sem fio e captura de movimentos.

The IMeasureU website features a dark-themed header with the company logo and navigation links for Markets, Products, Customer Wins, Media, Pricing, About Us, Academy, Support, Get in Touch, and Login. The main visual is a photograph of three male athletes in athletic gear. One is in the foreground, looking at a tablet, while two others are in the background. The central text reads: "Wearable sensor and software platform to measure, manage and maximize human movement". Below this, three callout boxes describe the platform's applications: "For tracking and rehabilitating elite and collegiate athletes" (with a "Capture.U" button), "For monitoring patients in and out of clinic" (with an "Orthopedics" button), and "For warfighter risk mitigation and long term monitoring" (with a "Defense" button). A "Start Scrolling" button is located at the bottom left of the main image area.

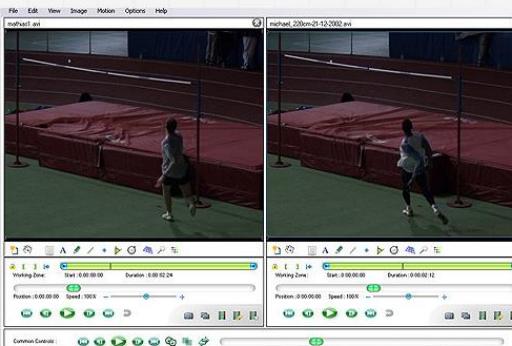
# COLETA DE DADOS E SENSORIAMENTO

 [KINOVEA](#)   [FEATURES](#)   [DOWNLOAD](#)   [HELP](#)   [FORUM](#)

## A MICROSCOPE FOR YOUR VIDEOS

Kinovea is a video annotation tool designed for sport analysis. It features utilities to capture, slow down, compare, annotate and measure motion in videos. ([Learn more](#)). Kinovea is completely free and open source.

[DOWNLOAD](#) **0.9.5**



 [OTHER DOWNLOADS](#)

 [hudl](#)   [Solutions](#) ▾   [Resources & Support](#) ▾   [Company](#) ▾    [Search Teams & Athletes](#)

## Change the way you see the game.

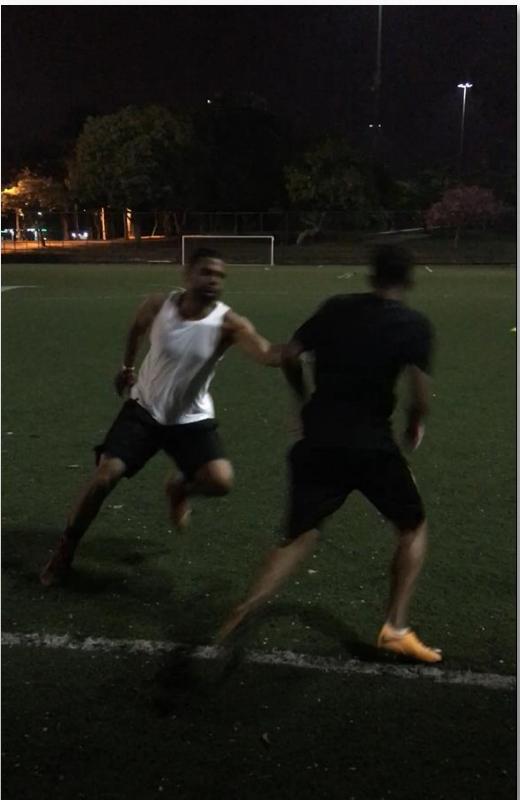
More than 230K teams across 40 sports use Hudl to capture, analyze and learn from video and data.

[Request a Demo](#)



[HIGH SCHOOLS](#) ▾   [CLUBS](#) ▾   [COLLEGES](#) ▾   [DIV I & PROFESSIONAL](#) ▾   [ATHLETES](#) ▾

# BASES DE DADOS



EAGLE-OPEN



FJERJ

## Campeonato Brasileiro de futebol

Campeonato Brasileiro de Futebol

[Data Card](#) [Code \(20\)](#) [Discussion \(7\)](#)

### About Dataset

#### Campeonato Brasileiro de Futebol

No total 8025 partidas de 2003 à 2022

#### Coleta dos dados

Coletamos os dados de páginas do Google

#### Validação dos dados

Criamos rotinas que validam as rodadas e a tabela do campeonato.

Montamos a tabela do campeonato calculando os pontos de cada partida.

#### Github do projeto

[https://github.com/adaoduque/Brasileirao\\_Dataset](https://github.com/adaoduque/Brasileirao_Dataset)

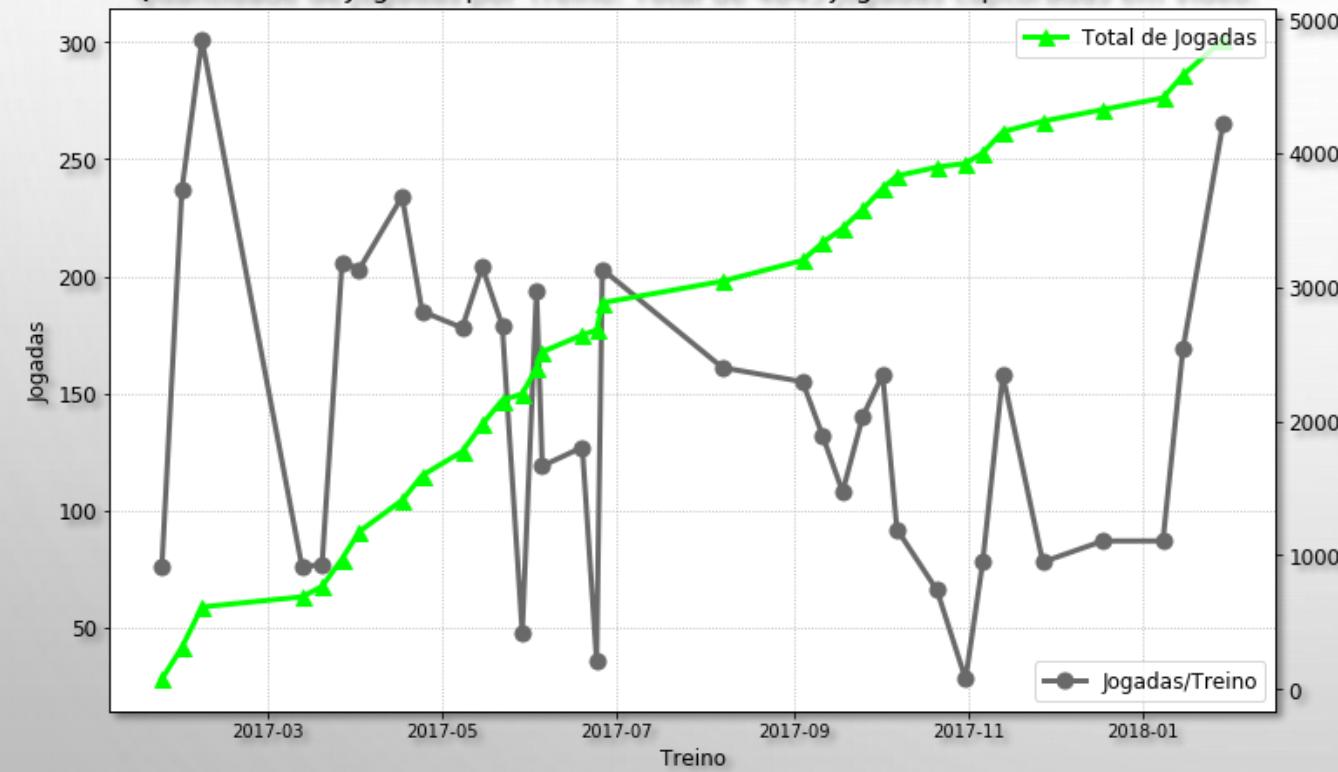
#### Contato

LinkedIn: [Meu LinkedIn](#)

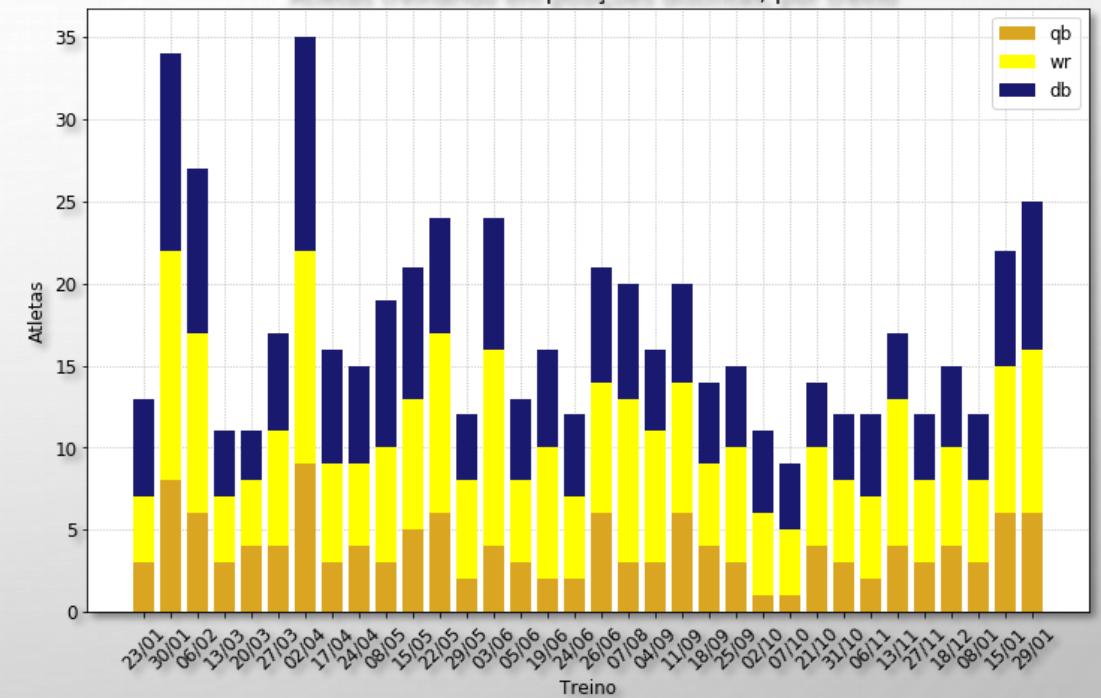
BRASILEIRÃO

# ESTATÍSTICA – VARIÁVEIS ALEATÓRIAS

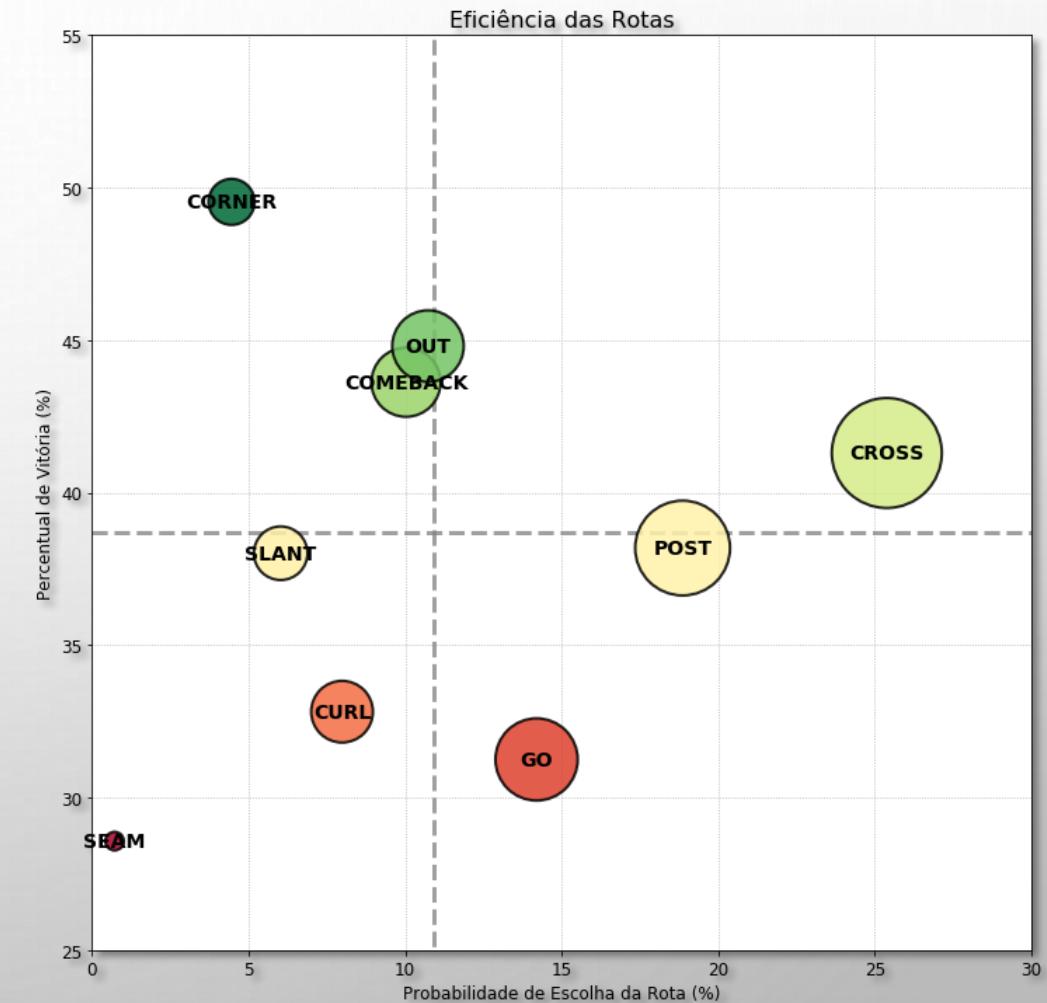
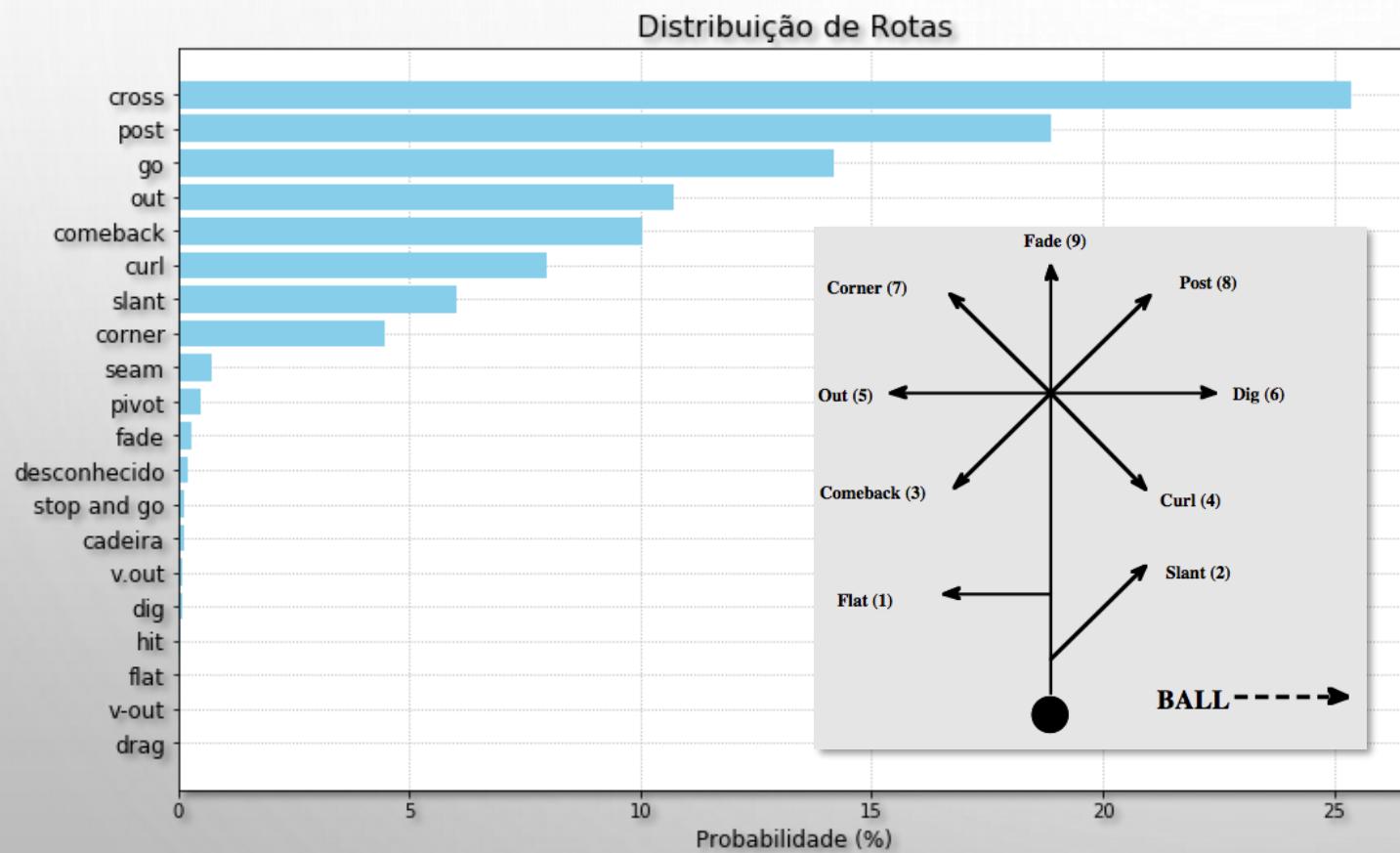
Quantidade de Jogadas por Treino. Total de 4845 jogadas capturadas em vídeo.



Atletas treinando em posições distintas, por treino



# ESTATÍSTICA – ANÁLISE EXPLORATÓRIA

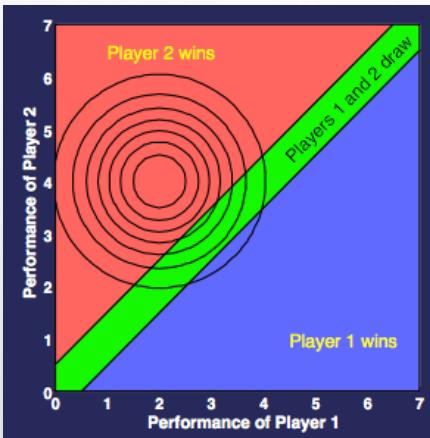


# ESTATÍSTICA PARA RANQUEAMENTO



ARPAD ELO

$$d_i \sim \mathcal{N}(\mu_i, \beta^2)$$



$$\delta_1 \sim \mathcal{N}(\mu_1 - \mu_2, \sqrt{2}\beta)$$

$$p_{ij} = E[\delta_i] = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{2}\beta}\right)$$

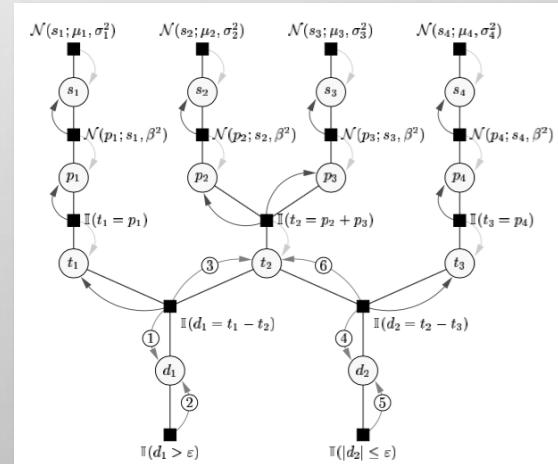
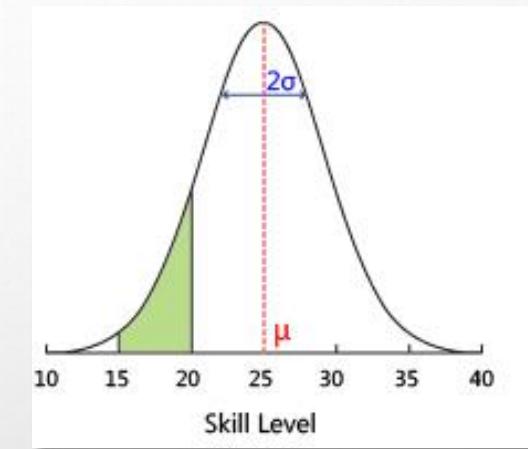


MARK GLICKMAN



TOM MINKA

$$p(\mu|\mathbf{y}, A) = \frac{P(\mathbf{y}|\mu, A)p(\mu)}{P(\mathbf{y}|A)}$$



# ESTATÍSTICA PARA RANQUEAMENTO

## World Football Elo Ratings

Ratings and Statistics as of Tuesday October 10 2023															
Rank	Team	Rating	Average		1 Year Change		Matches					Goals			
			Rank	Rating	Rank	Rating	Total	Home	Away	Neutral	Wins	Losses	Draws	For	Against
1	Argentina	2156	5	1985	+1	+15	1074	371	407	296	583	226	265	2033	1115
2	Brazil	2108	4	2000	-1	-61	1036	365	333	338	660	164	212	2251	923
3	France	2103	16	1786	+3	+98	903	458	332	113	451	264	188	1617	1249
4	Portugal	2022	19	1791	+3	+18	662	307	249	106	326	184	152	1148	748
5	Croatia	2016	12	1878	+8	+94	369	145	156	68	193	76	100	639	371
6	England	2011	4	1979	+8	+91	1120	492	509	119	654	210	256	2619	1095
7	Spain	2007	7	1942	-4	-38	747	326	292	129	436	137	174	1505	669
8	Belgium	1983	23	1752	-3	-42	843	403	362	78	375	295	173	1532	1328
9	Netherlands	1972	15	1845	-5	-68	864	409	348	107	444	233	187	1796	1106
10	Italy	1956	8	1915	-2	-37	868	390	316	162	461	170	237	1512	853

GO TO TOP		GO TO ME		Season One	World
Rank	CS Rating	Player	Wins	Win %	LB
1	38,051	chayJESUS ツ	244	93.85%	SA
2	33,057	JoJo	250	89.29%	NA
3	32,039	cadSaiyaJIN	203	92.69%	SA
4	32,000	Iaski	192	92.31%	NA
5	28,151	ALLE	143	92.86%	SA
6	27,448	unluckyL^^\n	135	92.47%	SA
7	26,798	Tweaky&	118	93.65%	NA
8	26,240	Cactus	172	88.21%	NA
9	26,239	filim	126	91.97%	NA
10	26,235	NINJABIN	176	89.80%	AS
11	26,000	elitelikethat	174	84.06%	NA
12	25,997	If you don't ligg-uh-luh-dang	145	88.98%	NA
13	25,572	A-	81	90.00%	EU
14	25,443	PERCY	98	84.48%	EU
15	25,435	red	133	85.26%	EU
16	25,203	Kristou	131	86.18%	EU
17	25,188	Goblin	95	87.96%	EU
18	25,02	rem1x	98	94.23%	SA
19	25,00	TeSeS	78	85.71%	EU
20	25,00	es3tag	120	83.33%	EU
21	25,00	AJ	80	86.02%	EU



**Carlsen, Magnus**

World Rank (Active):	1	B-Year:	1990
Federation:	Norway	Sex:	Male
FIDE ID:	1503014	FIDE title:	Grandmaster

**Rating**

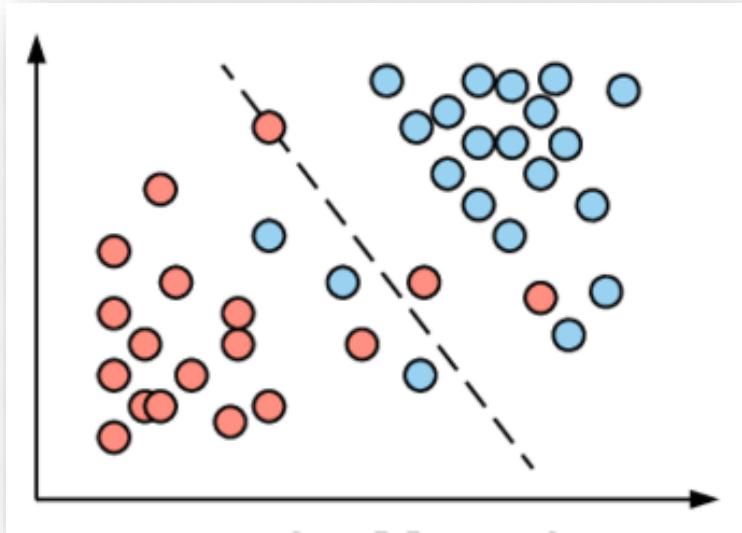
std 2839	rapid 2818	blitz 2887
----------	------------	------------

**SOK VS HAPPY**

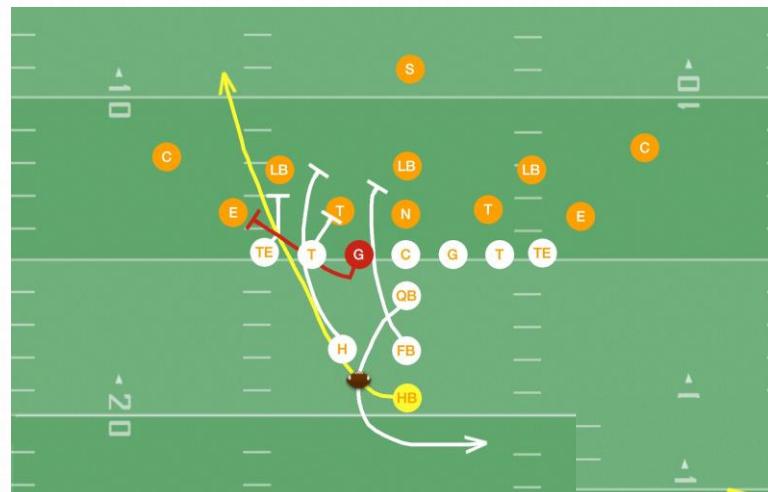
0 - 7

4 WINS VS. OPP	23
51% WR VS. ALL	81%
51% WR VS. RACE	87%
9th ELO RANKING	1st
1st RACE RANKING	1st

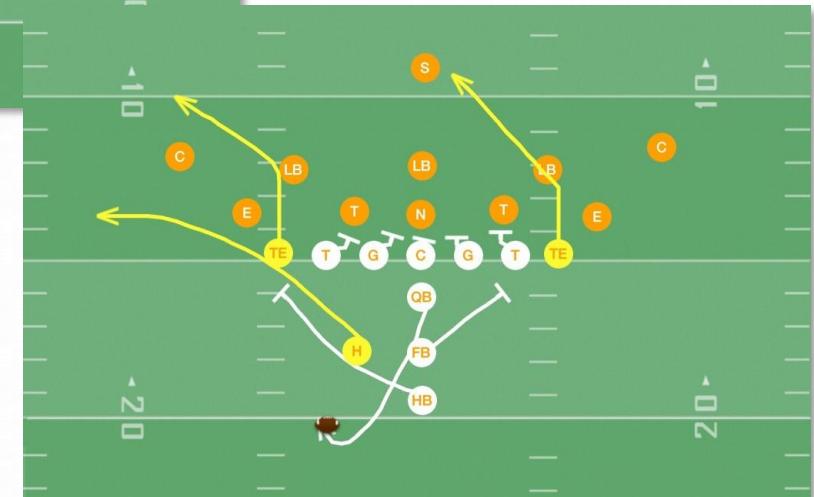
# MODELOS DE APRENDIZADO DE MÁQUINA: CLASSIFICAÇÃO



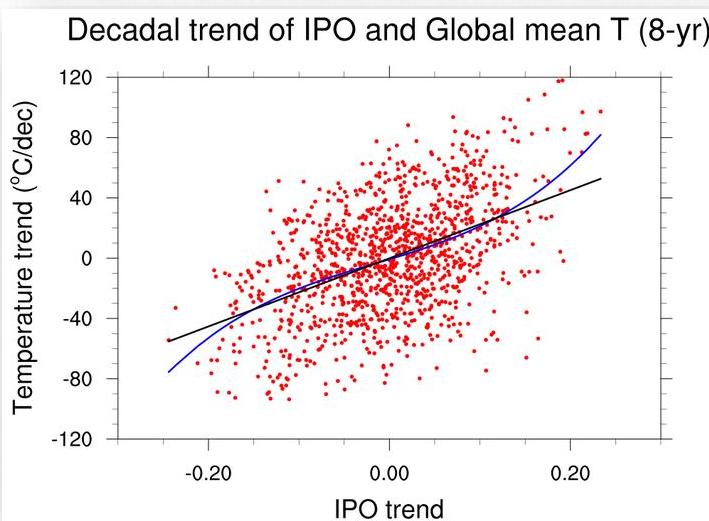
CLASSIFICAÇÃO



CORRIDA OU  
PASSE?



# MODELOS DE APRENDIZADO DE MÁQUINA: REGRESSÃO



REGRESSÃO



Ming Tang

Metallurgist (PhD) + Data Scientist

📍 Chicago, IL

✉️ Email

LinkedIn

Instagram

Github

Google Scholar

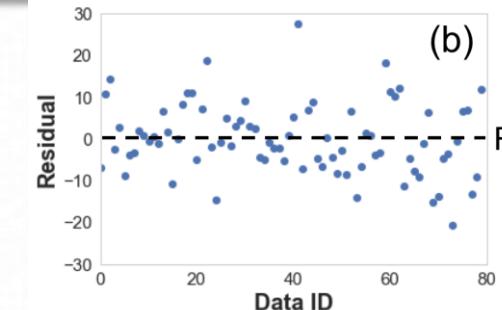
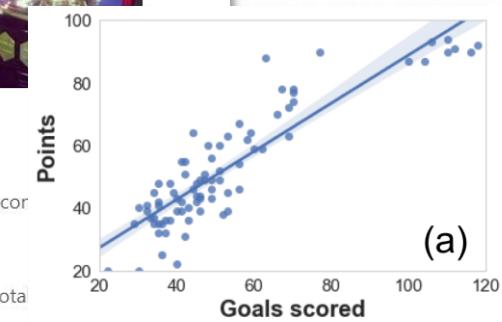
## Predicting Points In Soccer League By Regression



### Introduction

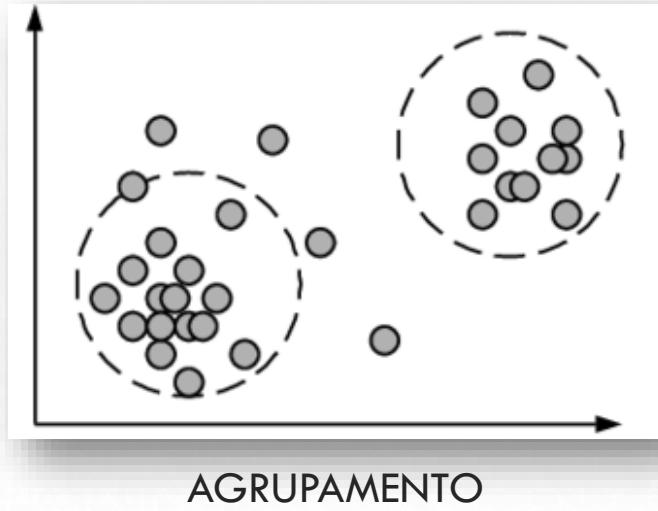
This blog is built on my second data science project at Metis/Chicago, which cor

- Scrape the information from the web
- Apply linear regression model to correlate features (goals) with target (total



QUANTOS PONTOS SERÃO  
FEITOS NO CAMPEONATO?

# MODELOS DE APRENDIZADO DE MÁQUINA: AGRUPAMENTO



[Home](#) > [Annals of Operations Research](#) > Article

**A robust method for clustering football players with mixed attributes**

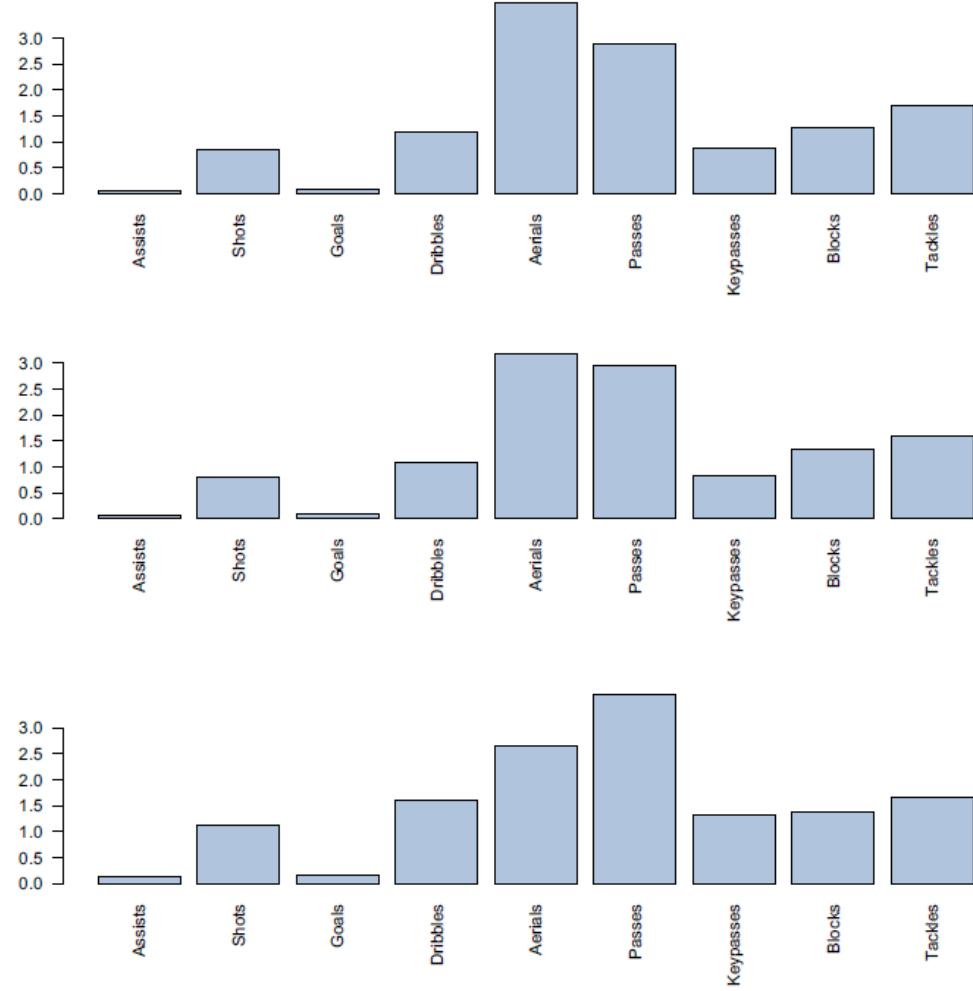
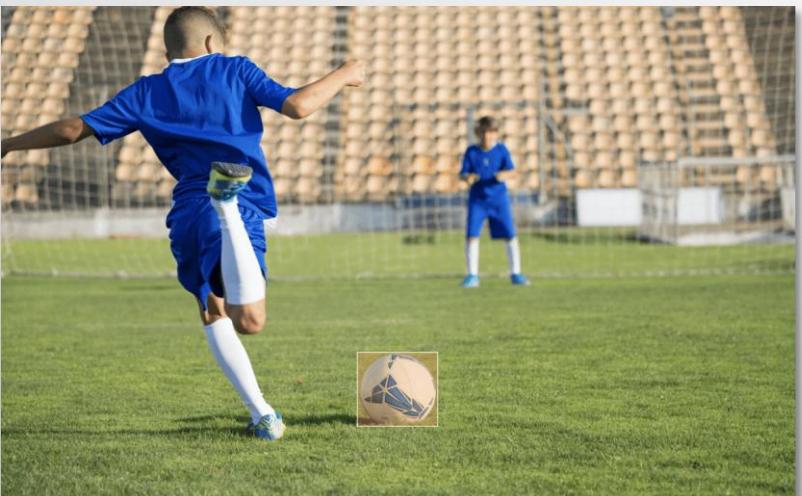
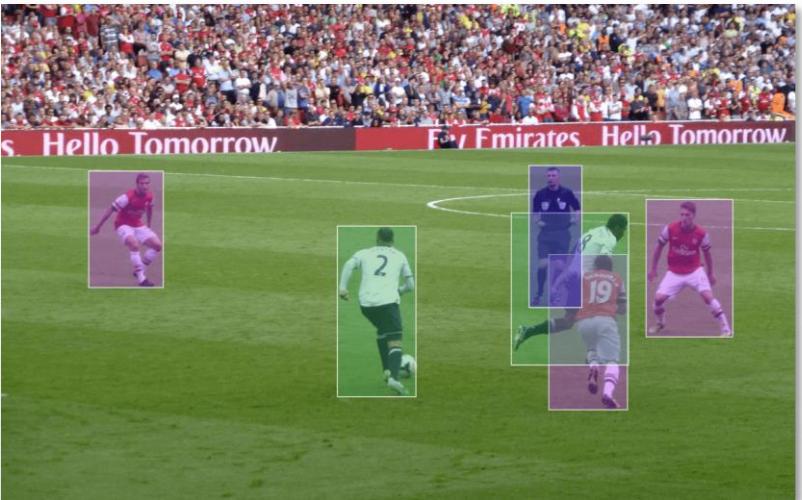


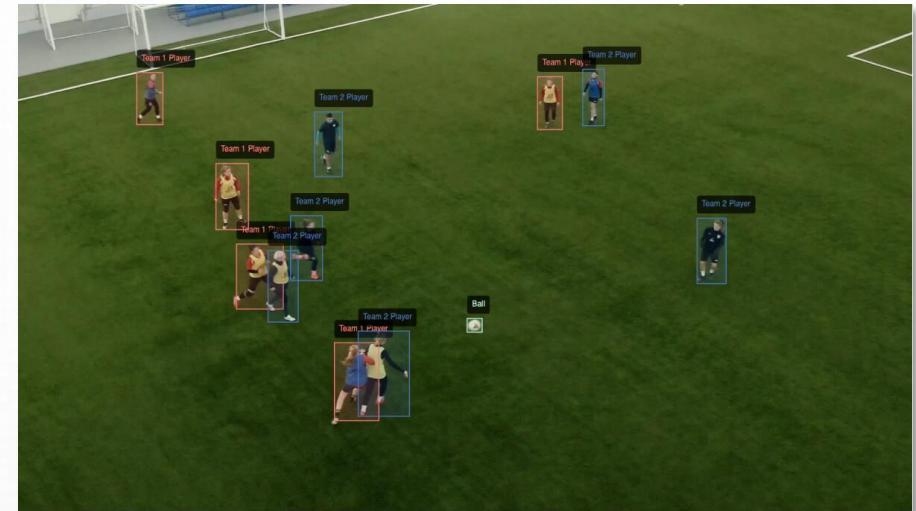
Fig. 5 Performance variables weighted with memberships within clusters. Cluster 1, 2, 3 top to bottom

QUAIS PERFIS DE ATLETAS DE FUTEBOL EXISTEM?

# VISÃO COMPUTACIONAL



OBJECT TRACKING



TRACKING & CLASSIFICATION



POSE DETECTION

# APLICAÇÕES

COMBATE



INDEPENDENTE



OBJETO

## BETTING ODDS EXPLAINED

Bookmakers

Bookmakers	1xbet	3.36	3.04	2.57
Palms Bet	Palms Bet	2.20 ▲	1.58 ▼	-
bet365	bet365	3.40 ▲	2.90 ▼	2.45 ▼
MARATHON	MarathonBet	3.35	2.93	2.43
BETWINNER	BetWinner	3.26 ▲	2.99 ▲	2.52 ▼
noxWIN	Noxwin	3.37	3.03	2.51

BETTING

# APRESENTAÇÃO DO PROFESSOR

# LINHA DO TEMPO: 2000'S...

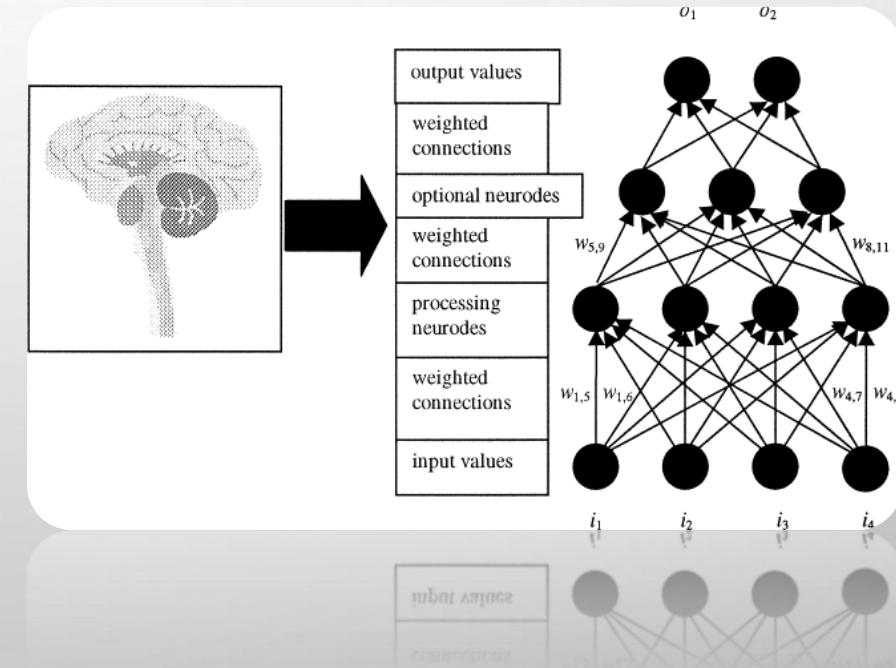
2000-2002

- **ESTUDANTE DO ENSINO MÉDIO E GAMER** DE AGE OF EMPIRES II
- DISPUTA DE CAMPEONATOS A NÍVEL INTERNACIONAL.
- **ESTATÍSTICA ELO** UTILIZADA PELA PLATAFORMA DA MICROSOFT PARA RANQUEAR OS JOGADORES.



2003-2008

- **ENGENHARIA ELETRÔNICA E COMPUTAÇÃO** NA UFRJ.
- **ATLETA E TÉCNICO DE FUTEBOL AMERICANO.**
- COMO FAZER UM ROBÔ PARA **ADIVINHAR AS JOGADAS DO MEU ADVERSÁRIO?** **REDES NEURAIS!**



# LINHA DO TEMPO: 2010'S...

2009-2013

- ANALISTA DE INTELIGÊNCIA DE NEGÓCIOS JR, TRABALHANDO COM MANUTENÇÃO DE CAMINHÕES DE MINERAÇÃO NO MESTRADO.
- PUBLICAÇÃO DE UM ARTIGO DE PREVISÃO DE JOGADAS DO PAYTON MANNING EM UM CONGRESSO NACIONAL.
- ATLETA E TÉCNICO DE FUTEBOL AMERICANO NO FLAMENGO E NO VASCO.



2014-2018

- CIDADÃO CIENTISTA DE DADOS, TRABALHANDO E DESENVOLVENDO A ÁREA DE CIÊNCIA DE DADOS EM VÁRIAS EMPRESAS, COMO CIENTISTA DE DADOS CONSULTOR.
- OPORTUNIDADE COMO OLHEIRO DA SELEÇÃO BRASILEIRA ME LEVA A DESENVOLVER UM ALGORITMO DE RANQUEAMENTO DE ATLETAS INSPIRADO NO ELO.

EAGLE SKILLS WR



Olá, Philippe

RANK 3 +1  
CA\$H 360.77 +53.73  
SKILL 89 +13

↗ Comeback  
➡ Go

Resumo do Treino

34 REP	19 WIN	56% PV	+53.73 CA\$H
--------	--------	--------	--------------

Rotas Coberturas

	Slant	3	10.96	3.65	Press	15	32.14	2.14
	Cross	7	10.15	1.45	square			
	Comeback	5	7.86	1.57	Press	8	9.05	1.13
	Fade	3	6.83	2.28				
	Out	3	5.45	1.82	Off in	10	8.21	0.82
	Post	6	5.16	0.86	Off			
	Corner	4	5.07	1.27	square	1	4.32	4.32
	Pivot	2	3.00	1.50				
	Curl	1	-0.76	-0.76				

Spn	1	-0.18	-0.18
Yard	5	3'00	J20
Convex	1	2'03	J51
Down	0	2'18	0'90
Off	3	2'42	1'85
End	3	9'83	5'58

# LINHA DO TEMPO: 2020'S...

2019-2023

- TRANSIÇÃO PARA O **MUNDO DO EMPREENDEDORISMO** ABRINDO UMA **CONSULTORIA DE DATA SCIENCE** COM DOIS AMIGOS.
- **INCUBAÇÃO** DO PROJETO DE DOUTORADO DE RANQUEAMENTO DE ATLETAS DO VASCO NA INCUBADORA DA COPPE / UFRJ.



Hoje

- SERVIÇO DE ANÁLISE DE DESEMPENHO PARA O VASCO ALMIRANTES
- SERVIÇO DE ANÁLISE DE DESEMPENHO B2C PARA ATLETAS CARIOCAS DE FUTEBOL AMERICANO.
- SERVIÇO DE SCOUTING DE CAMPEONATO DESENVOLVIDO PARA A BRAZILIAN FOOTBALL LEAGUE (BFA)
- SERVIÇO DE SCOUTING DE TREINO DESENVOLVIDO PARA O INSTITUTO REAÇÃO (PILOTO)
- GRAVAÇÃO, TRANSMISSÃO E SCOUTING PARA A FEDERAÇÃO CARIOLA DE JUDO.
- APlicativo de GESTÃO DE TREINO PARA ESCOLINHAS DE FUTEVOLLEY





**Diego Rodrigues**

Lead Data Scientist at Interagente & Eagle Analytics  
Rio de Janeiro, Rio de Janeiro, Brazil  
1K followers · 500+ connections

 See your mutual connections

[Join to view profile](#)

 Interagente

 Universidade Federal do Rio de Janeiro

## About

Fifteen years of professional and academic involvement in business intelligence, predictive analysis, machine learning systems, and data mining. Development of Data Science projects in various sectors of the economy: e-commerce, management consulting, predictive maintenance, and Sports Analytics. Ph.D. in Computational Intelligence from COPPE/UFRJ.

Currently, I lead the Data Science department at Interagente and Eagle Sports Analytics, a company incubated at the UFRJ technology park.

At Interagente, our goal is to enhance our clients' business efficiency through the use of artificial intelligence. Our algorithms can be used to gain insights, support, or automate critical decisions. Contact us to discover how to use AI to improve your company's results. [www.interagente.ai](http://www.interagente.ai)

At EAGLE, we use data science algorithms and artificial intelligence to measure and report skills for athletes in different sports. Contact us to learn more about the solutions we are developing for the sports market.

# DIEGO RODRIGUES DSC

- Doutor em Inteligência Computacional COPPE / UFRJ
- 15 anos de experiência com ciência de dados aplicada em diferentes áreas.
- Lead Data Scientist na Interagente e na Eagle Sports Analytics.
- Leciono: Análise Exploratória de Dados & Redes Neurais e Deep Learning com Tensorflow.

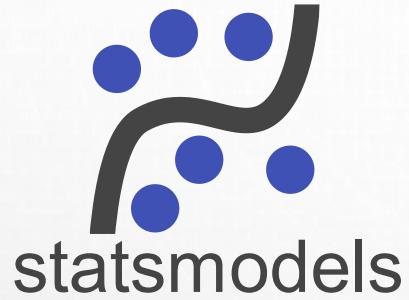
# TÓPICOS EM DATA SCIENCE PARA O ESPORTE

- NIVELAR OS ALUNOS EM **CONCEITOS ESTATÍSTICOS** E DE **APRENDIZADO DE MÁQUINA**.
- PREPARAR OS **CIENTISTAS DE DADOS ESPORTIVOS** PARA OS **DESAFIOS DO DIA-A-DIA**.
- DESENVOLVER **ABORDAGENS INOVADORAS** DO USO DE DADOS PARA DESENVOLVER **APLICAÇÕES** PARA O ESPORTE.
- **PUBLICAR ARTIGOS** E SE TORNAR **REFERÊNCIA CIENTÍFICA** NA ÁREA DE CIÊNCIA DE DADOS ESPORTIVO.



# **ORGANIZAÇÃO DO AMBIENTE**

# ORGANIZAÇÃO DO AMBIENTE



# APRESENTAÇÃO DA TURMA