

footure
academy

ESPECIALIZAÇÃO

CIÊNCIA NOVO
DE **DADOS**

NO FUTEBOL

Aula 4 - Tipos de dados

Caio Batatinha

AGENDA

- Análise qualitativa x análise quantitativa
- Tipos de dados
- Dados de súmula
- Dados de evento
- Dados de rastreamento
- Dados abertos

OBJETIVOS DE APRENDIZAGEM

- Conhecer os diferentes tipos de dados utilizados na análise esportiva
- Entender as diferenças na coleta e respectiva aplicação

REVISÃO CONCEITOS AULA PASSADA

- Tipicamente existem quatro tipos de profissionais diferentes num departamento de Analytics: Gerente de dados, cientista de dados, analista de dados e engenheiro de dados.
- O departamento pode operar de forma independente como um centro de pesquisa ou de forma integrada, alocada outro departamento específico.
- Alternativamente, o modelo de consultoria podia ser adotada para acelerar processos.

ANÁLISE QUALITATIVA E QUANTITATIVA

QUALITATIVA

Foco: Aspectos intangíveis, interpretações.

Dados: Descritivos

Perguntas motivadoras: Como e por quê?

Vantagens: Compreensão melhor de contextos. Captura atributos fundamentais como aspecto mental do atleta.

Desvantagens: Velocidade de processamento. Subjetividade que pode levar a vieses.

QUANTITATIVA

Foco: Aspectos mensuráveis

Dados: Numéricos

Perguntas motivadoras: O quê, quantos e qual a frequência?

Vantagens: Objetividade e escalabilidade. Permite comparações não enviesadas.

Desvantagens: Limitado aos fatores mensuráveis, perda de contexto na análise.

ANÁLISE QUALITATIVA

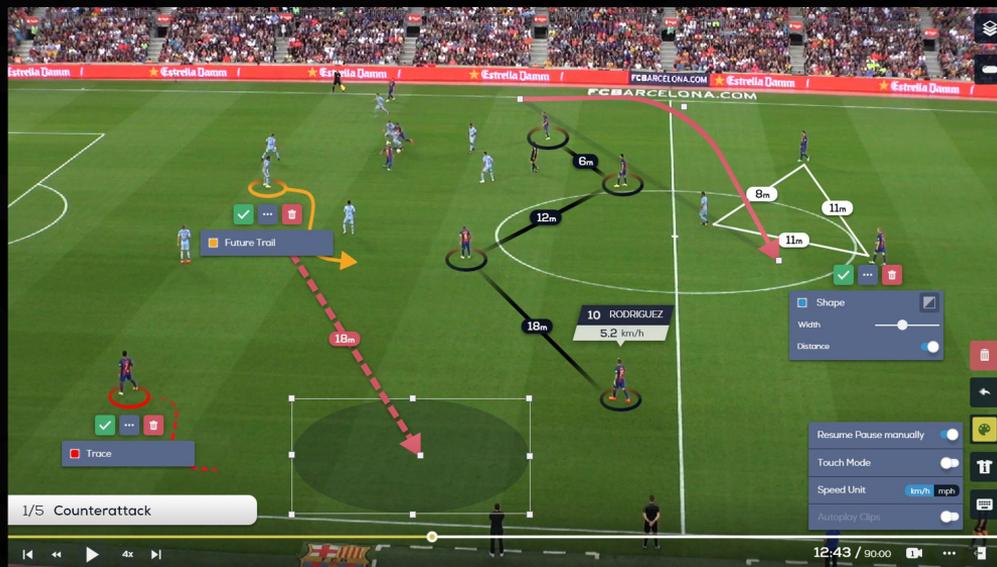
ANÁLISE DE VÍDEO

A análise de vídeo se fundamenta em aspectos mais qualitativos do jogo.

Se o objeto de estudo é uma equipe adversária, coletam-se observações para tentar responder questões como:

- Como essa equipe ataca?
- Por quê utiliza essa saída de bola?
- Como é a transição ofensiva e defensiva?
- Como é a pressão utilizada?
- Porquê esses jogadores foram escalados nessas funções?

A partir dessas observações, um relatório é produzido para descrever as suas idéias de jogo.



ANÁLISE QUANTITATIVA

RELATÓRIO PÓS JOGO

Estatísticas pós jogo é um exemplo de uma análise quantitativa rápida, onde podemos responder levantar hipóteses sobre a performance das equipes envolvidas a partir de perguntas como:

- Quem teve mais controle das ações?
- Qual levou mais perigo ao gol adversário?
- Quem ocupou mais a área adversária?
- Quem foi mais efetivo nos duelos?



O PAPEL DO PROFISSIONAL DE DADOS

O Profissional da ciência de dados no futebol tem como foco e tarefa principal realizar análises **qualitativas**.

O escopo da análise depende intrinsecamente da seleção dos tipos de dados para a execução da tarefa.

Os tipos de dados variam em:

- Granularidade
- Forma de coleta
- Aplicação
- Limitação e potenciais erros

TIPOS DE DADOS

**DADOS DE
SÚMULA**

**DADOS DE
EVENTO (EVENT
DATA)**

**DADOS DE
RASTREAMENTO
(TRACKING
DATA)**

**DADOS
FINANCEIROS**

**DADOS
AGREGADOS**

DADOS FÍSICOS

DADOS DE SÚMULA

Descreve os eventos mais importantes ocorridos numa partida.

Granularidade:

- Gols, cartões, minutos jogados, número de faltas.

Coleta :

- Normalmente coletado manualmente pelos observadores da partida.

Aplicação:

- Análises simples de partidas. Análises do histórico de jogadores.

Limitações:

- Baixo nível de granularidade, faltam detalhes e incapaz de fornecer contexto devido para cada jogador.

The screenshot shows the 'fotlink' website interface. At the top, there's a search bar and navigation icons. The main header identifies the player as Gabrielle Jordao Portilho (GJ), Brazilian, 29 years old, and her current club as Corinthians. Below this, there are tabs for 'Visão Geral', 'Anotações', 'Avaliações', 'Carreira', 'Desempenho' (selected), 'Jogos', and 'Relatórios PRO'. The 'Desempenho' section includes a filter area with dropdowns for 'Temporada: Todas', 'Equipe: Todas', and 'Competição: Todas', along with a 'Limpar Filtros' button and an 'Agrupar por:' dropdown set to 'Nenhum critério'. Below the filters is a 'Visão Geral' table with columns for 'Temporada', 'Nome da Competição', 'Clube', and various performance metrics represented by icons.

Temporada	Nome da Competição	Clube	🕒	📊	👕	🏠	⚽	🟡	🔴	🏆
2024	CBF - Supercopa Futebol Fe...	Corinthians	256'	3	3	0	1	1	0	0
2024	CBF - Brasileiro Feminino A1...	Corinthians	1.197	16	15	4	1	0	0	0
2023	CBF - Brasileiro Feminino A1...	Corinthians	1.213'	16	13	3	3	0	0	1
2020	CBF - Brasileiro Feminino (C...	Corinthians	383'	9	3	7	2	0	0	0

DADOS DE TRANSFERÊNCIA

Descreve aspectos de natureza financeira de um jogador ou clube.

Granularidade:

- Data de transferência, clubes envolvidos, transferências passadas, tipo de transferência, taxas envolvidas, valor de mercado.

Coleta :

- Publicado pelos próprios clubes ou agentes. Base de dados da liga ou ferramentas de análise de mercado.

Aplicação:

- Valuation de jogadores, planejamento de orçamento, análise de mercado e tendências de transferências.

Limitações:

- Fontes não uniformes. Falta de detalhamento disponibilidade para todas transferências, influência externa de agentes e marcas.

#	Player	Age	Season	Nat.	Left	Joined	Fee
1	Vitor Roque Centre-Forward	18	23/24	BR	Athletico-PR Série A	Barcelona LaLiga	€30.00m
2	Lucas Beraldo Centre-Back	20	23/24	BR	São Paulo Série A	Paris SG Ligue 1	€20.00m
3	Gabriel Moscardo Defensive Midfield	18	23/24	BR	Corinthians Série A	Paris SG Ligue 1	€20.00m
4	Matheus Franca Attacking Midfield	19	23/24	BR	Flamengo Série A	Crystal Palace Premier League	€20.00m
5	Luiz Henrique Right Winger	23	23/24	BR	Real Betis LaLiga	Botafogo Série A	€16.00m
6	Deivid Washington Centre-Forward	18	23/24	BR	Santos Série A	Chelsea Premier League	€16.00m
7	Ângelo Right Winger	18	23/24	BR	Santos Série A	Chelsea Premier League	€15.00m
8	Artur Right Winger	25	23/24	BR	Palmeiras Série A	Zenit S-Pb Premier Liga	€15.00m
9	Nicolás de la Cruz Central Midfield	26	23/24	AR	River Plate Liga Profesional	Flamengo Série A	€14.50m
10	Murillo Centre-Back	21	23/24	BR	Corinthians Série A	Nottm Forest Premier League	€12.00m
11	Marlon Gomes Central Midfield	20	23/24	BR	Vasco da Gama Série A	Shakhtar D. Premier Liga	€12.00m
12	Kevin Left Winger	21	23/24	BR	Palmeiras Série A	Shakhtar D. Premier Liga	€12.00m
13	Natan Centre-Back	22	23/24	BR	RB Bragantino Série A	Napoli Série A	€12.00m

DADOS DE EVENTO (EVENT DATA)

Descreve todas ações com bola de uma partida executada pelo jogador.

Granularidade:

- Ações como desarmes, passes, chutes com o tempo, posição e resultado da ação.

Coleta :

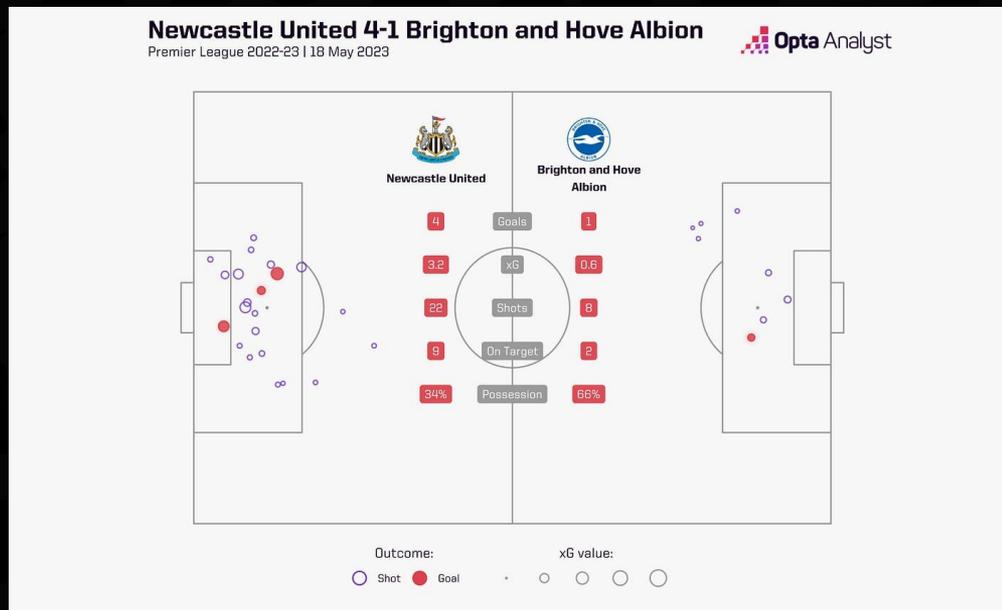
- Ferramentas automáticas de captura a partir do feed de transmissão ou manual por profissionais treinados pelas fornecedoras de dado.

Aplicação:

- Análise de ações individuais, análise de desempenho, identificação de padrões e contribuição de jogadores.

Limitações:

- Falta de contexto tática e não captura ações e eventos ocorridos sem bola.



DADOS AGREGADOS

Descreve ações cumulativas

Granularidade:

- Estatística totais para ações capturadas pelos dados de evento.

Coleta :

- Deriva da agregação dos dados de evento.

Aplicação:

- Análises comparativas. Criação de indicadores e métricas para análise de desempenho, scouting,

Limitações:

- Contexto e tipicamente é utilizado de forma absoluta para descrever índice de qualidade sem combinar outras análises.

Campeonato Brasileiro Série A History 2024 Série A Overview Scores & Fixtures Squad & Player Stats Nationalities Other 2024 Leagues Back to top

Player Advanced Goalkeeping 2024 Série A Hide non-qualifiers for rate stats [Glossary](#) [Toggle Per90 Stats](#)

Rk	Player	Nation	Pos	Squad	Age	Born	90s	Goals				Expected			Launched			Passes			Goal Kicks			Crosses					
								GA	PKA	FK	CK	OG	PSxG	PSxG/90	PSxG+/-	/90	Cmp	Att	Cmp%	Att (GK)	Thr	Launch%	AvgLen	Att	Launch%	AvgLen	Opp	Stp	Stp%
1	Jhoel Ricardo Riedi	BRA	GK	Fortaleza	36-069	1988	33.0	32	5	0	6	0	43.8	0.24	+11.8	+0.36	169	453	37.3	611	128	49.9	39.7	263	56.3	44.1	578	40	6
2	Fábia	BRA	GK	Fluminense	44-045	1980	32.0	34	4	0	5	0	45.7	0.29	+11.7	+0.37	73	252	29.0	567	132	32.3	30.5	238	29.0	23.2	470	26	5
3	Wêverton	BRA	GK	Palmeiras	36-337	1987	32.0	25	1	2	3	1	35.2	0.27	+11.2	+0.35	132	424	31.1	792	152	42.2	38.1	205	43.9	40.6	374	35	9
4	John	BRA	GK	Botafogo (RJ)	28-275	1996	30.0	22	5	0	1	0	31.2	0.25	+9.2	+0.31	112	297	37.7	817	110	26.2	30.5	186	44.6	38.5	398	23	5
5	Cleiton Schwenhobar	BRA	GK	Red Bull Bragantino	27-087	1997	25.0	29	0	0	3	1	36.6	0.29	+8.6	+0.34	113	342	33.0	650	121	44.6	37.9	179	29.1	33.0	379	24	6
6	Cássio	BRA	GK	Cruzeiro	37-161	1987	17.0	15	1	1	2	0	21.5	0.28	+6.5	+0.38	39	103	37.9	353	83	23.8	31.0	116	16.4	22.4	249	9	3
7	Gabriel	BRA	GK	Juventude	32-048	1992	28.5	42	8	1	2	0	48.1	0.27	+6.1	+0.21	105	281	37.4	743	118	27.6	32.1	208	36.5	34.6	402	35	8
8	Hugo Souza	BRA	GK	Corinthians	25-288	1999	17.0	19	4	0	2	0	24.5	0.26	+5.5	+0.32	58	170	34.1	344	87	36.0	33.8	124	37.1	34.5	253	14	5
9	Léo Jardim	BRA	GK	Vasco da Gama	29-239	1995	33.0	49	0	1	10	0	54.3	0.31	+5.3	+0.16	163	405	40.2	618	139	45.0	37.2	275	46.2	40.4	535	29	5
10	Aquatin Rossi	ARG	GK	Flamengo	29-085	1995	31.0	34	2	0	7	1	38.1	0.33	+5.1	+0.16	129	317	40.7	928	134	27.9	31.3	170	34.1	33.1	386	39	10
11	Aquatin Marchesin	ARG	GK	Grêmio	36-243	1988	24.0	28	6	0	3	1	31.9	0.22	+4.9	+0.21	119	350	34.0	503	74	49.1	38.8	191	53.9	41.3	383	23	6
12	Matheus Mendes	BRA	GK	Atlético Mineiro	25-249	1999	11.1	19	1	0	2	0	23.8	0.37	+4.8	+0.43	31	75	41.3	237	50	20.3	26.2	77	35.1	27.1	156	9	5
13	Gustavo	BRA	GK	Criciúma	31-249	1993	28.3	41	3	4	3	1	44.6	0.28	+4.6	+0.16	145	335	43.3	511	105	44.2	37.4	253	43.1	36.4	477	21	4
14	Walter	BRA	GK	Cuiabá	36-362	1987	32.0	41	6	1	6	0	45.4	0.27	+4.4	+0.14	179	399	44.9	636	123	38.4	35.3	279	55.6	42.3	591	24	4
15	Bento	BRA	GK	Ath Paranaense	25-157	1999	7.0	4	0	0	1	0	6.5	0.24	+2.5	+0.36	45	86	52.3	123	18	39.8	34.8	47	78.7	55.7	86	6	7
16	Serouj Rochet	URU	GK	Internacional	31-236	1993	23.0	19	3	0	0	0	21.2	0.23	+2.2	+0.10	98	276	35.5	473	78	45.7	36.7	120	50.0	39.3	277	29	10
17	Carlos Miquel	BRA	GK	Corinthians	26-036	1998	6.0	7	1	0	0	2	7.1	0.21	+2.1	+0.35	33	66	50.0	139	26	29.5	32.5	46	54.3	43.8	98	4	4
18	Anderson	BRA	GK	Cruzeiro	26-254	1998	16.0	20	1	0	2	1	20.8	0.31	+1.8	+0.11	28	92	30.4	311	52	21.2	28.9	122	21.3	26.3	235	11	4
19	Anthony	BRA	GK	Internacional	22-296	2002	2.0	1	0	0	0	0	1.8	0.18	+0.8	+0.41	14	45	31.1	46	5	52.2	38.6	26	80.8	56.2	47	1	2
20	Fabrizio	BRA	GK	Red Bull Bragantino	24-152	2000	1.0	2	0	0	0	0	2.7	0.40	+0.7	+0.69	3	7	42.9	16	7	37.5	33.1	5	20.0	20.2	12	0	0
21	Cássio	BRA	GK	Corinthians	37-161	1987	3.0	3	0	0	0	0	3.7	0.27	+0.7	+0.24	11	25	44.0	68	16	26.5	31.6	19	36.8	38.0	29	3	0
22	Muriel Becker	BRA	GK	Vitória	37-274	1987	1.0	0	0	0	0	0	0.6	0.32	+0.6	+0.61	11	21	52.4	20	4	60.0	43.9	10	90.0	60.0	19	1	5
23	Matheus Pesianato	BRA	GK	Cuiabá	32-139	1992	1.0	0	0	0	0	0	0.5	0.18	+0.5	+0.53	7	15	46.7	15	4	60.0	32.0	11	81.8	51.5	14	0	0
24	Rafael Cabral	BRA	GK	Grêmio	34-178	1990	5.1	8	1	0	0	0	8.4	0.32	+0.4	+0.06	25	73	34.2	106	20	44.3	36.7	47	55.3	40.7	71	5	7
25	Gabriel Delfim	BRA	GK	Atlético Mineiro	22-127	2002	1.0	1	0	0	0	0	1.1	0.37	+0.1	+0.11	1	5	20.0	21	7	19.0	25.6	9	11.1	23.2	10	2	0
26	Alisson	BRA	GK	Criciúma	29-159	1995	4.7	7	0	0	1	0	7.1	0.28	+0.1	+0.02	22	74	29.7	98	22	46.9	38.3	54	51.9	43.4	84	5	6
27	Caíque	BRA	GK	Grêmio	27-106	1997	3.9	6	0	0	1	1	5.1	0.28	+0.1	+0.02	13	45	28.9	56	17	55.4	40.8	24	58.3	45.5	50	3	6

DADOS DE RASTREAMENTO (TRACKING DATA)

Descreve ações contínuas em uma partida de futebol.

Granularidade:

- Estatísticas alta granularidade com coordenadas cartesianas de todos os jogadores frame a frame.

Coleta :

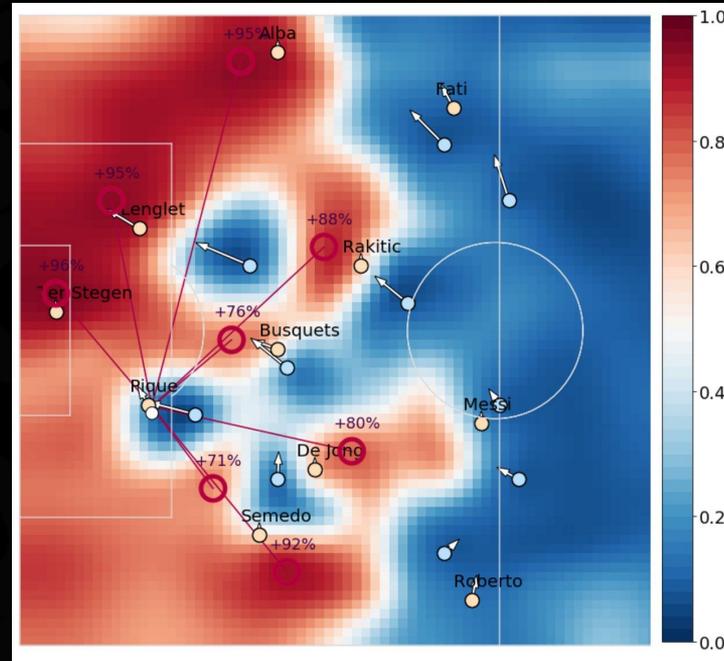
- Através de um sistema multicâmeras.

Aplicação:

- Modelos matemáticos espaciais, análise tática, movimentação de jogadores com e sem a bola, estruturas de pressão.

Limitações:

- Necessita de uma alta capacidade armazenamento e qualidade na captura. Apresenta ruídos e necessita de um cuidado maior no tratamento de dados.



DADOS FÍSICOS

Descreve as capacidades físicas de um atleta

Granularidade:

- Estatísticas como Frequência cardíaca, aceleração, distância percorrida, carga metabólica.

Coleta :

- Através de coletes e outros aparelhos de rastreamento via GPS.

Aplicação:

- Monitoramento de condicionamento físico, carga de treinos, desenvolvimento atlético, prevenção de lesões.

Limitações:

- Dados podem ser influenciados por variáveis externas como temperatura, escolhas táticas, altitude e utilizada como medida direta de performance sem contexto.

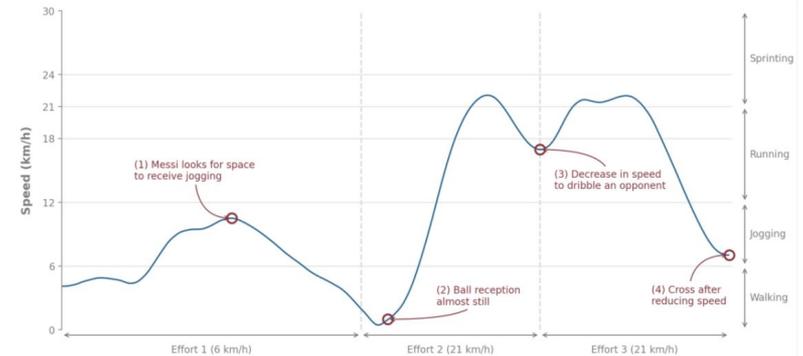


Fig. 1. Above, the evolution of Messi's speed in a game situation. The first highlighted moment represents a movement to approach (1) the ball. Then, the action is formed on (2) the pass reception, (3) a ball drive with a dribble, and (4) a cross. Below, the video frames of the four highlighted instants.

**O que vem
por aí**

POSE DATA

Descreve ações biomecânicas dos jogadores

Granularidade:

- Estatísticas ultra detalhadas como orientação corporal, execução física.

Coleta :

- Tecnologias avançadas de captura de movimento e algoritmo de estimativa de orientação corporal.

Aplicação:

- Análise de performance física, ajustes técnicos motores, prevenção de lesões.

Limitações:

- Alto custo de tecnologia, complexidade e dificuldade de interpretação.

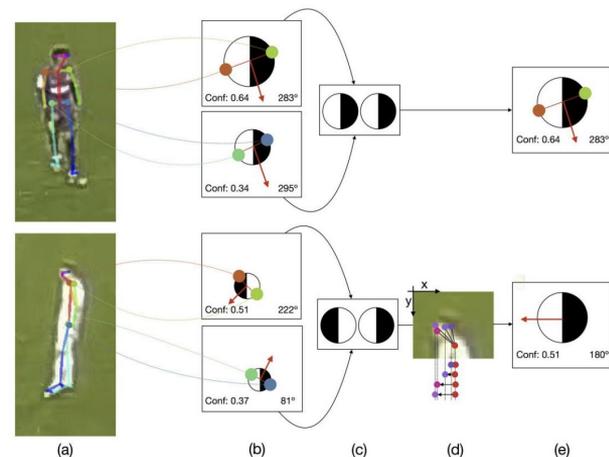


Figure 4. Steps to estimate the pose orientation of a given player and its LR-side parameter: (a) OpenPose output, (b) coordinates belonging to shoulders and hip are mapped in a 2D field, individual LR-side and orientation values are obtained. (c) A simple check is performed to see if both LR-parameters coincide; if they do not and their confidence difference is below a certain threshold, (d) the face direction is checked by comparing the X coordinate of all face parts with the neck; (e) a final estimation is obtained according to: (top) higher confidence value, or (bottom) face direction.

DADOS NEUROLÓGICOS

Descreve a atividade mental e comportamentos de um atleta.

Granularidade:

- Estatísticas como tempo de reação, velocidade de tomada de decisão, nível de concentração e estresse.

Coleta :

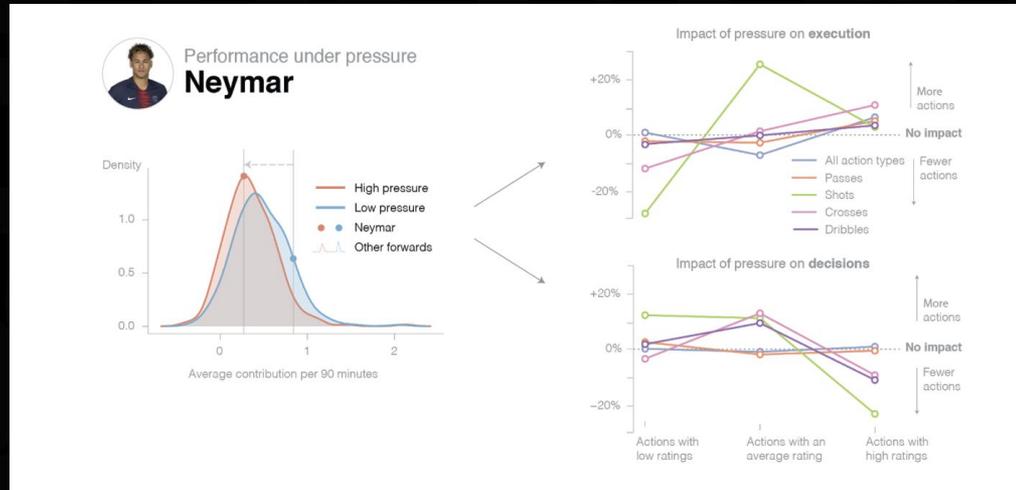
- Através de interfaces computador-cérebro, monitores de frequência cardíaca, monitores ópticos, óculos de realidade virtual.

Aplicação:

- Treinamento cognitivo, monitoramento de situações de alta pressão, prevenção e monitoramento da saúde mental.

Limitações:

- Alto nível de tecnologia e expertise para analisar dados de natureza mais objetiva.



TIPO	SÚMULA	TRANSFERÊNCIAS	EVENTO	AGREGADOS	RASTREAMENTO	FÍSICOS	POSE	NEUROLÓGICOS
GRANULARIDADE	Baixa	Baixa	Média	Média	Alta	Alta	Alta	Alta
COLETA	Manual	Via clube e ferramentas de mercado	Ferramentas automatizadas	Via dados de evento	Sistema multicâmeras	Aparelhos de monitoramento via GPS	Tecnologias de captura de movimento	Monitores ópticos, frequência cardíaca, óculos VR
APLICAÇÃO	Análises rápidas, histórico de atleta	Valuation, planejamento de orçamento	Análise de desempenho, identificação de padrões	Criação de KPIs para scouting e desempenho	Análises táticas e avançadas de movimentação	Monitoramento de carga, prevenção de lesões	Ajustes técnicos motores, análise física.	Treinamento cognitivo, situações de alta pressão.
LIMITAÇÃO	Baixo nível de detalhe	Falta de uniformidade de dados	Não captura ações sem bola	Falta de contexto	Muitos ruídos, dificuldade de armazenamento	Influência de variáveis externas	Alto custo de captação	Alto custo e dificuldade de interpretação

DADOS ABERTOS



Dúvidas?

REFERÊNCIAS UTILIZADAS

<https://static1.squarespace.com/static/5b048119f2e6b103db959419/t/5dcef4b8d980ba7a8814e619/1573844340461/Head%2C+Shoulders%2C+Hip+and+Ball...+Hip+and+Ball%21+Using+Pose+Data+to+Leverage+Football+Player+Orientation+by+Arbues-Sanguesa%2C+A%2C+Haro%2C+G%2C+Ballester%2C+C+%26+Martin%2C+A+>

<https://www.sofascore.com/>

<https://www.transfermarkt.com/>

<https://statsbomb.com/what-we-do/soccer-data/360-2/>

<https://fbref.com/en/comps/24/Serie-A-Stats>

<https://football-data.co.uk/>

<https://understat.com/>

<https://github.com/statsbomb/open-data>

<https://github.com/fivethirtyeight/data/tree/master/soccer-spi>

https://people.cs.kuleuven.be/~pieter.robberchts/repo/bransen-ssac19-Choke_or_Shine.pdf

<https://barcainnovationhub.fcbarcelona.com/investigation/is-it-worth-the-effort-understanding-and-contextualizing-physical-metrics-in-soccer/>

OBRIGADO!



@CcBatatinha



[linkedin.com/in/caio-batatinha/](https://www.linkedin.com/in/caio-batatinha/)



caio.batatinha@footure.com.br
caiocbatatinha@gmail.com