



Olá! Antes de começarmos quero fazer um teste estatístico... Quem aqui gosta de futebol, levanta a mão por favor. Certo, desses que gostam de futebol, quem acompanha seu time assistindo os jogos? Legal, e agora quem vai até o estádio assistir os jogos? Muito bom, e agora desses, quem é corintiano? (Comentar o resultado, se muitos corintianos dizer que era o resultado esperado).



Vamos lá, primeiro, parabéns! Parabéns por terem vindo em uma sexta feira aprender e compartilhar um pouco do seu conhecimento e experiência, vocês fazem eventos como o Devcamp se tornarem muito bons e produtivos.



Eu sou o Luiz, desenvolvedor no Corinthians, faço parte da equipe do CIFUT, o Centro de Inteligência do Futebol. A equipe é formada por seis analistas de desempenho, que acompanham e analisam muita informação sobre praticamente tudo no futebol, eles são realmente muito bons no que fazem, minha visão sobre o futebol mudou completamente e aposto que, se qualquer um aqui parar para conversar 10 minutos com um deles, vai ver que futebol é muito mais complexo do que podemos imaginar. O meu trabalho lá é ajudar a simplificar a captura e o processamento das informações para gerar o conhecimento que é transmitido para a comissão e para os jogadores.



Gooool!!

O MOMENTO
MAIS ESPERADO
DO JOGO

GOOOOOOL!!!

O gol, o momento que todos esperam aflitos, com o coração na mão, e quando acontece parece que todo o peso do mundo sai das suas costas, naqueles segundos suas preocupações desaparecem e uma alegria toma seu coração. Mas por que? Por que o gol tem esse efeito sobre nós?

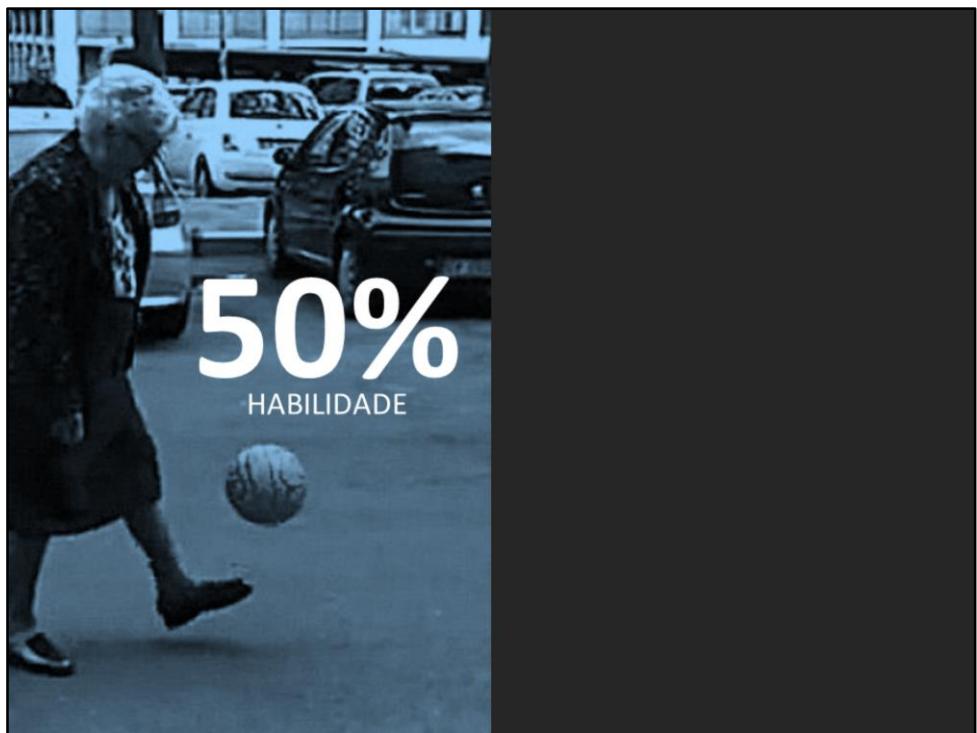


O MOMENTO
MAIS ESPERADO
DO JOGO

EXTREMAMENTE RARO

GOOOOOOL!!!

Simples, por que ele é um evento extremamente raro, tão simples mas muito difícil de conquistar, é essa raridade que faz o futebol ser o esporte mais praticado no mundo e que enche estádios, todos querem ver a raridade acontecer diante de seus olhos.



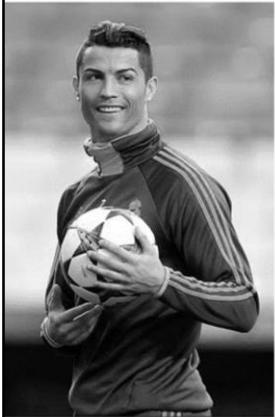
No livro “Os números do jogo” é mostrada uma estatística que diz que os gols acontecem na metade das vezes graças a habilidade dos jogadores...



... e na outra metade por sorte. Quem acha que isso é verdade?

The World's Most Powerful Celebrities

[Log in](#) | [Sign up](#) | [Conn...](#)



#30 Cristiano Ronaldo

[+ Follow](#) (321)

2014 Celebrity 100 Earnings

\$80 Million

Athlete, Soccer



Next #31 >

Cristiano Ronaldo on
Forbes Lists

#30 Celebrity 100 (2014)

#14 in Money

#2 in Press

Age	30
Source Of Wealth	Soccer
Salary/Winnings	\$52 M
Endorsements	\$28 M
Residence	Madrid, Spain
Citizenship	Portugal
Marital Status	Single
Children	1
Agent	Luis Correia
Agency	Gestifute International

<http://j.mp/zerando-jogo-01>

Infelizmente parece ser, por isso as grandes equipes investem milhões para contratar os melhores jogadores...

The World's Most Powerful Celebrities

[Log in](#) | [Sign up](#) | [Conn...](#)



#30 Cristiano Ronaldo

[+ Follow](#) (321)

2014 Celebrity 100 Earnings

\$80 Million

Athlete, Soccer

Age	30
Source Of Wealth	Soccer
Salary/Winnings	\$52 M
Endorsements	\$28 M
Residence	Madrid, Spain
Citizenship	Portugal
Marital Status	Single
Children	1
Agent	Luis Correia
Agency	Gestifute International

Cristiano Ronaldo on
Forbes Lists

#30 Celebrity 100 (2014)

#14 in Money

#2 in Press

[Next #31 >](#)

**\$ MILHÕES
EM HABILIDADE**

<http://j.mp/zerando-jogo-01>

... para garantir que os seus 50% de habilidade sejam suficientes para superar os 50% de habilidade dos adversários.



Mas investir milhões não é a melhor opção, ainda mais se você está em um país onde um dólar vale mais que um café com pão na chapa.



Então, para encontrar uma opção melhor, precisamos entender o jogo, desmembrar toda a sua complexidade. A melhor forma de fazer isso é usando a teoria dos jogos.



Quem já assistiu “Uma mente brilhante”?

Excelente filme, conta a história do matemático John Nash, daqui a pouco a gente vai falar um pouco mais sobre ele, o filme ajudou a popularizar a teoria dos jogos.

“

O estudo de modelos matemáticos de conflito e cooperação entre tomadores de decisão com inteligência racional.

Myerson, Roger B. (1991). Game Theory: Analysis of Conflict, Harvard University Press, p. 1. Chapter-preview links, pp. vii–xi.

<http://j.mp/zerando-jogo-02>

A teoria dos jogos é um ramo da matemática que estuda as situações estratégicas entre jogadores, considera as cooperações e concorrências e tem o objetivo de ajudar a entender as estratégias que podem trazer o melhor retorno para os jogadores.



Vou mostrar um exemplo da teoria dos jogos dentro do futebol



...o pênalti. Na forma mais simples, são dois jogadores interagindo, ambos têm que tomar uma decisão aleatória no momento que o juiz apita.

GOLEIRO			
BATEDOR	ESQUERDA	CENTRO	DIREITA
ESQUERDA			
CENTRO			
DIREITA			

O goleiro escolhe um lado para defender, o batedor escolhe um lado para chutar.

		GOLEIRO		
		ESQUERDA	CENTRO	DIREITA
BATEDOR	ESQUERDA		GOL	GOL
	CENTRO	GOL		GOL
	DIREITA	GOL	GOL	

Se o goleiro escolhe a esquerda e o batedor escolhe o centro, então o goleiro perde e o batedor ganha.

		GOLEIRO		
		ESQUERDA	CENTRO	DIREITA
BATEDOR	ESQUERDA	DEFESA	GOL	GOL
	CENTRO	GOL	DEFESA	GOL
	DIREITA	GOL	GOL	DEFESA

Já se o goleiro escolhe a direita e o batedor também escolhe a direita, então, teoricamente, o batedor perde e o goleiro ganha.

Bem simples né?



E se o goleiro souber em qual lado o batedor provavelmente vai chutar antes do juiz apitar? A decisão não vai ser mais aleatória e sim estratégica. O mesmo vale para o batedor, se ele souber para qual lado o goleiro provavelmente vai pular a habilidade dele é potencializada e as chances de fazer o gol são maiores. Conseguir essa informação pode definir a equipe campeã.



NA VIDA REAL
NÃO É TÃO
SIMPLES

Só saber o que o adversário provavelmente vai fazer não é garantia, ele pode saber que você tem essa informação e então, mudar de estratégia, agora temos um problema, e nosso amigo John Nash vai nos ajudar.

RESPOSTAS COM TEORIA DOS JOGOS

Aplicando os conceitos de teoria de jogos, para cada situação temos três respostas...

RESPOSTAS COM TEORIA DOS JOGOS

**MELHOR
OPÇÃO**

Maior Probabilidade

A melhor possível, que considera que o adversário não sabe que você tem as informações sobre ele.

RESPOSTAS COM TEORIA DOS JOGOS

**MELHOR
OPÇÃO**

Maior Probabilidade

**PIOR
OPÇÃO**

Menor Probabilidade

A pior possível, que é a ação que você tem que evitar ao máximo

RESPOSTAS COM TEORIA DOS JOGOS

**MELHOR
OPÇÃO**

Maior Probabilidade

**OPÇÃO
EQUILIBRADA**

Equilíbrio de Nash

**PIOR
OPÇÃO**

Menor Probabilidade

E a opção equilibrada, que na teoria dos jogos seria o equilíbrio de Nash, aqui encontramos a melhor escolha considerando que o adversário também vai procurar a melhor escolha para ele.

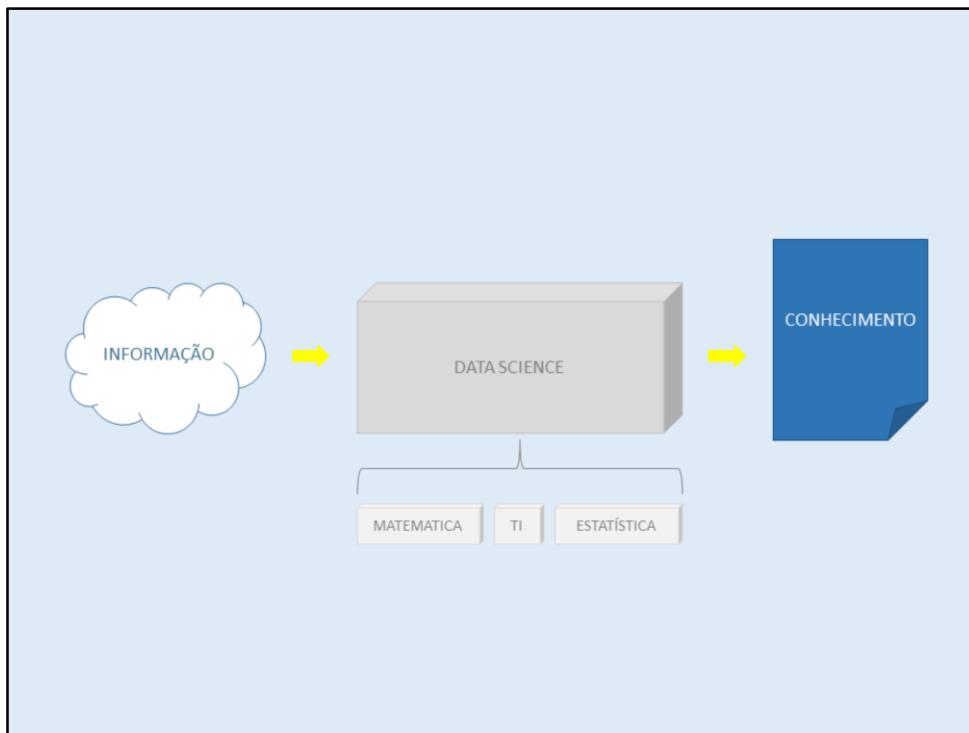
**PARA ENTENDER O JOGO
PRECISAMOS DE**

**MUITA
INFORMAÇÃO**

No exemplo do pênalti, como é um jogo simples, jogadores habilidosos conseguem encontrar a opção equilibrada por instinto, só com um pouco de informação sobre o adversário. Mas para a maioria dos casos precisamos de informação, muita informação, aí entra a Data Science.



Mas o que é Data Science, ou ciência dos dados? É Big Data? Não, é muito mais que isso.



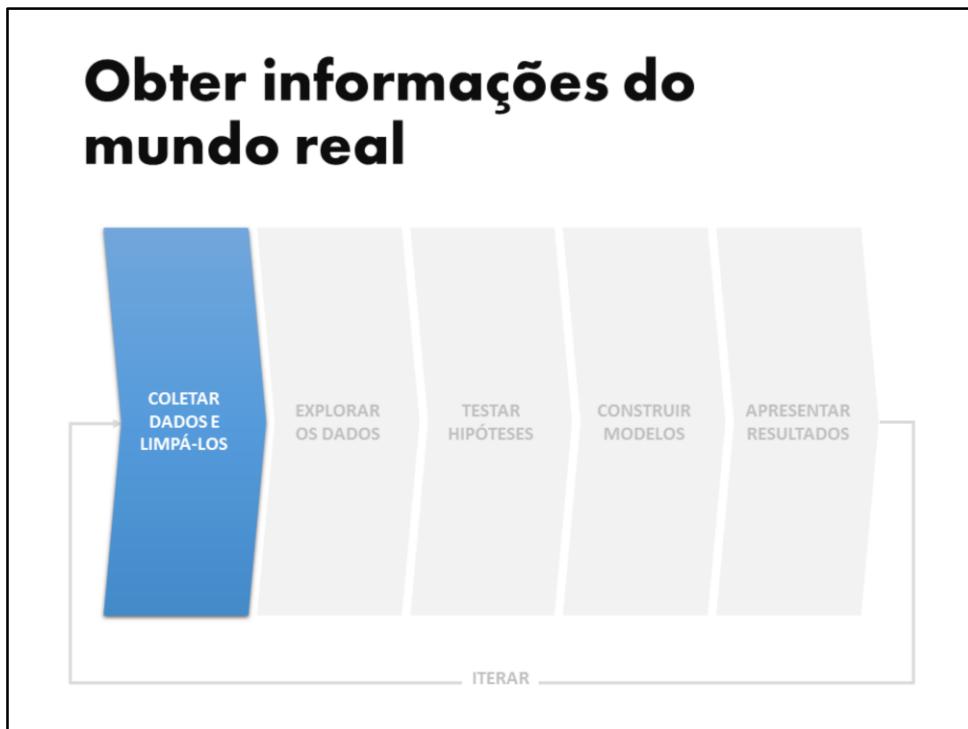
De forma genérica, a Data Science, pode ser definida como a extração de conhecimento da informação, utilizando a metodologia científica. Engloba, além da tecnologia da informação, a matemática e estatística. A Data Science existe a muito tempo e é usada em muitas áreas, por exemplo na biologia para entender a evolução, na medicina para entender e prever epidemias e atualmente no comércio para recomendar os produtos que você provavelmente vai gostar mais.

O Ciclo da Data Science



A Data Science tem um fluxo básico para seguir, que serve de guia para criar seus projetos de ciência de dados.

Obter informações do mundo real



Inicialmente você precisa descobrir formas de capturar os dados do mundo real para que você possa analisar depois, os dados podem vir de qualquer jeito, de múltiplos sistemas e ambientes, então você também precisa trata-los para conseguir trabalhar com eles.

Encontrar perguntas para responder



Quando você tem dados organizados e limpos, você pode separar uma amostra e procurar por padrões de comportamento e relacionamentos, nessa fase você começa a formar hipóteses sobre o mundo real, usando como base seu conhecimento sobre o que está estudando e as pistas que os dados te apresentam.

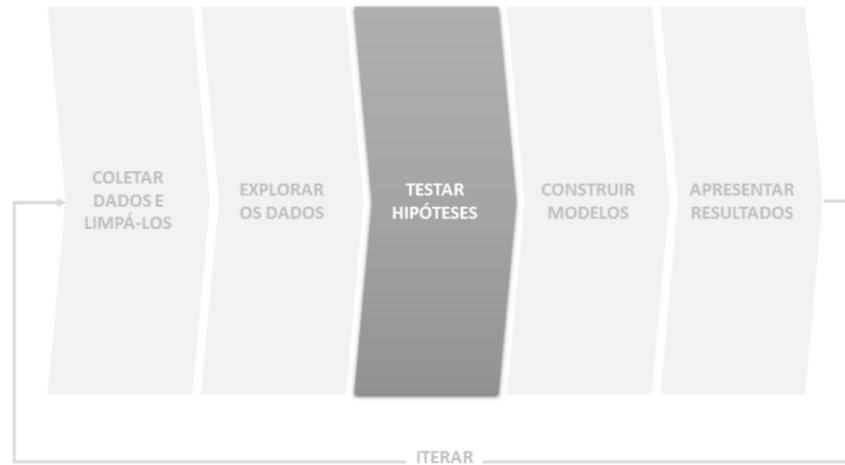
Testar as respostas das perguntas



Com suas hipóteses formadas chegou a hora de experimentá-las, usando a maior concentração de dados disponível você consegue confirmar ou descartar suas perguntas e formular novas, nessa fase o volume de dados vai te dar mais precisão nos resultados, então quanto mais big melhor.

Testar as respostas das perguntas

A computação e a internet permitem trabalhar com mais informações



Essa é a diferença da Data Science para os estudos científicos tradicionais, a computação e a internet te permitem trabalhar com muito mais informações, espalhadas pelo mundo todo, de forma rápida e eficiente.

Preparar para reprodução



Com seus modelos testados e confirmados você não precisa mais ficar refazendo todos os testes manualmente, nesse passo podemos construir modelos reproduzíveis e ensinar as máquinas a fazer o trabalho matemático pesado pela gente.

Criar produtos de dados



O produto final pode ser um dashboard com gráficos interativos, um relatório impresso ou até uma planilha no Excel.

Criar produtos de dados

Transmitir o conhecimento de forma correta



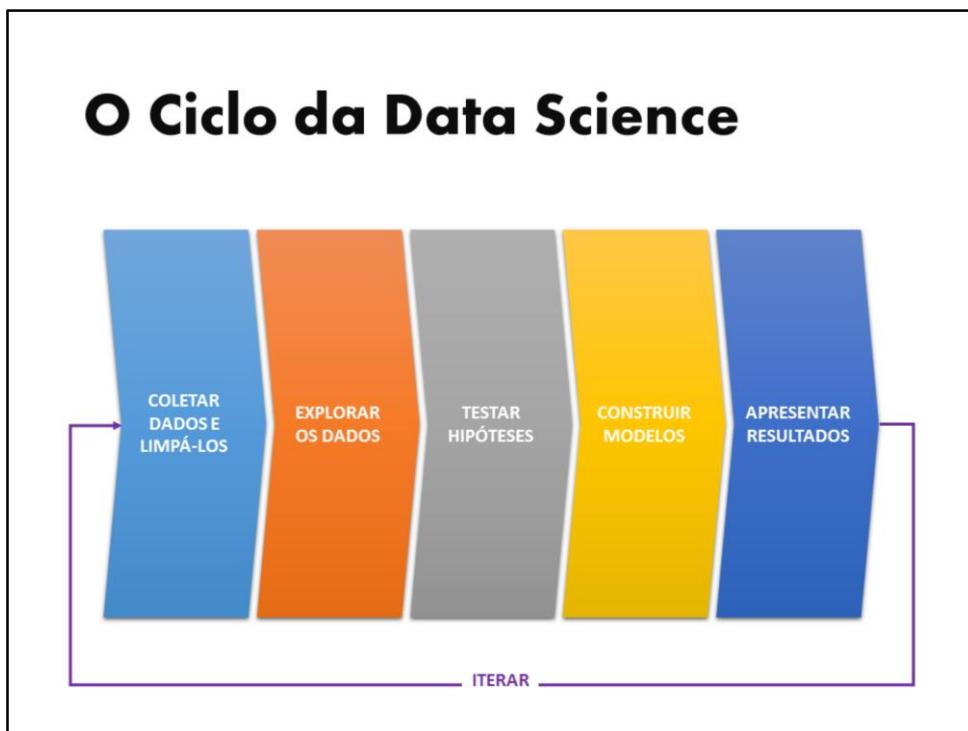
Talvez esse seja o passo mais importante, se não soubermos transmitir o conhecimento produzido, todos os passos anteriores terão sido em vão, você pode ter a rede de computadores mais poderosa do mundo, capaz de processar terabytes de dados por segundo, mas se a visualização do resultado passar a mensagem errada, todo o seu estudo estará comprometido.

Evoluir com o mundo real



Como tudo no universo, os dados estão em constante evolução, é necessário reanalisar os dados constantemente para formular novas hipóteses e encontrar novas pistas que te colocarão a frente de todos os concorrentes, sempre alguma coisa passa desapercebido e pode ser detectada numa segunda, terceira ou milésima passada de olhos.

O Ciclo da Data Science

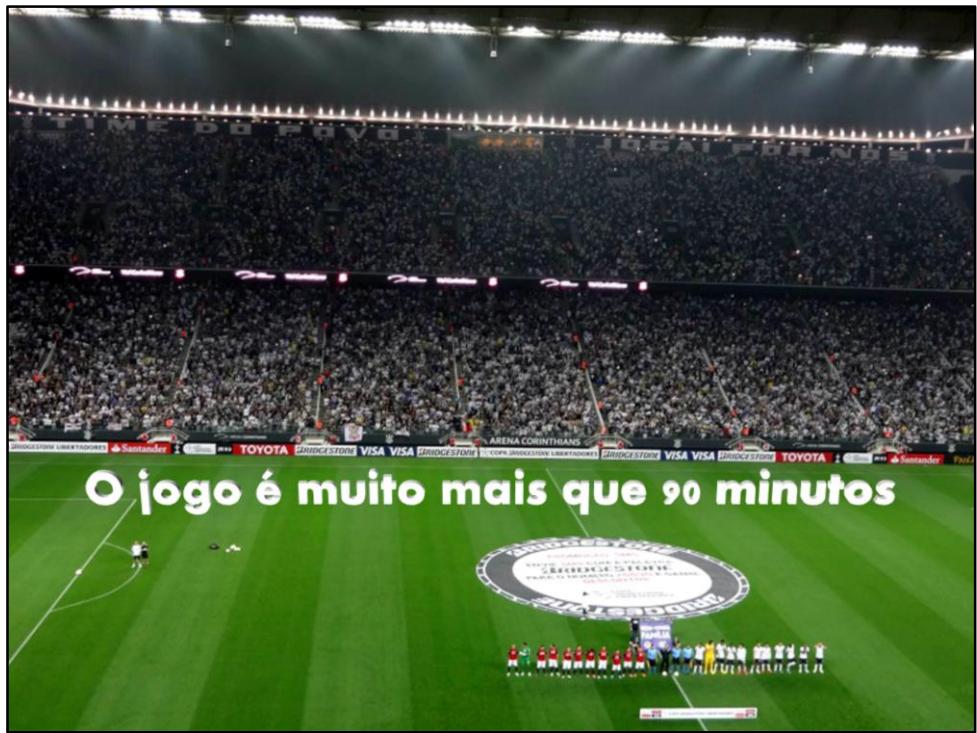


Dominar cada passo desse ciclo é dominar a Data Science, acredito que ainda não conseguimos usar todo o poder que essa área tem para oferecer em nenhum ramo de estudo atual.



Agora vou mostrar um pouco de como funciona o CIFUT.

Quem acha que o jogo se resume aos minutos entre o apito de início e de final de jogo?

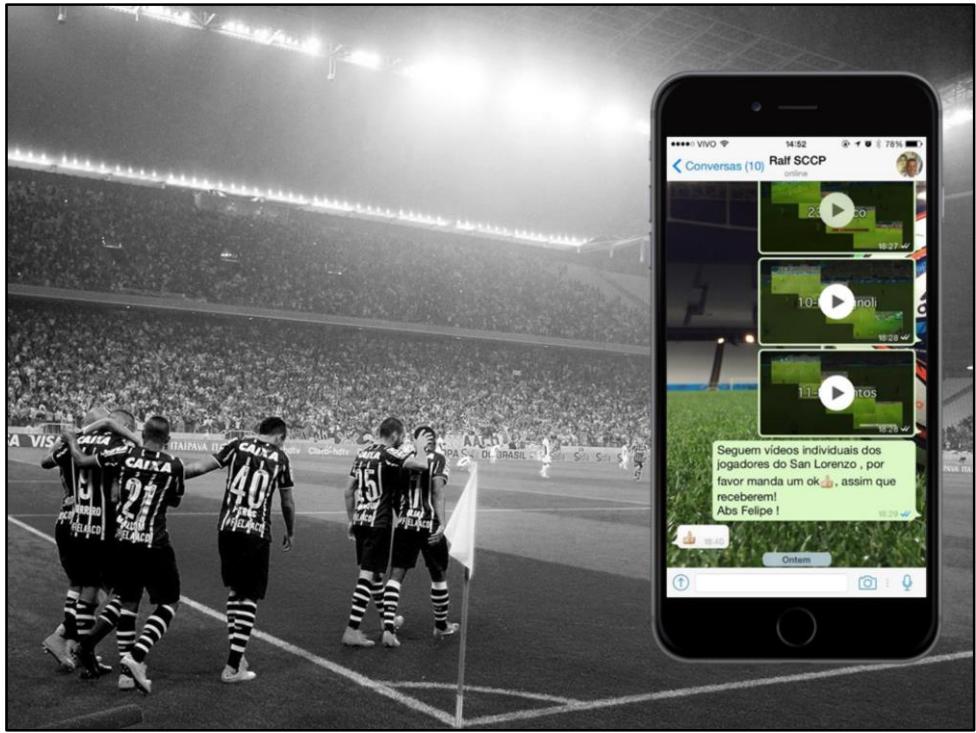


Não, é muito mais, cada jogo tem uma rotina exaustiva que começa...



Captura de Dados Pré Jogo

...Alguns dias antes do jogo, a equipe analisa o passado da equipe adversária, seus pontos fortes, suas fraquezas. Quando necessário analisa profundamente um jogador, assistindo horas e mais horas de jogo atrás de jogo, até definir um padrão comportamental do adversário.



Todas as informações são repassadas para a comissão técnica e para os jogadores antes do jogo, para que conheçam os possíveis problemas e oportunidades.

Captura de Dados Durante o Jogo



Durante o jogo a equipe coleta muita informação, utilizando iPads e fazendo marcações em tempo real sobre tudo no jogo. Todos os passes, toques na bola e até comportamentos dos jogadores longe da bola é registrado.

Apresentação: Durante o Jogo



Ainda no jogo algumas informações são apresentadas para a comissão, informações que ajudam o Tite a se preparar para o segundo tempo e dão uma visão geral do que está acontecendo. Infelizmente, o regulamento dos jogos não permite que essa informação vá em tempo real para eles na beira do campo e acredito que está certo, pensando em outras equipes menores o acesso a informação em tempo real seria uma vantagem injusta.



Após o jogo o trabalho continua, a análise feita durante o jogo é revista e aprofundada, os erros e acertos da equipe são separados para que a comissão trate com cada jogador individualmente.

O objetivo é entender o que aconteceu de certo e errado nos mínimos detalhes.

Apresentação: Antes e após o Jogo

As informações coletadas antes, durante e depois do jogo são armazenadas, analisadas e viram relatórios, vários relatórios. Alguns on-line, outros impressos, com o objetivo de levar a informação o mais simples, direta e corretamente possível para quem precisa.

Apresentação: Antes e após o Jogo



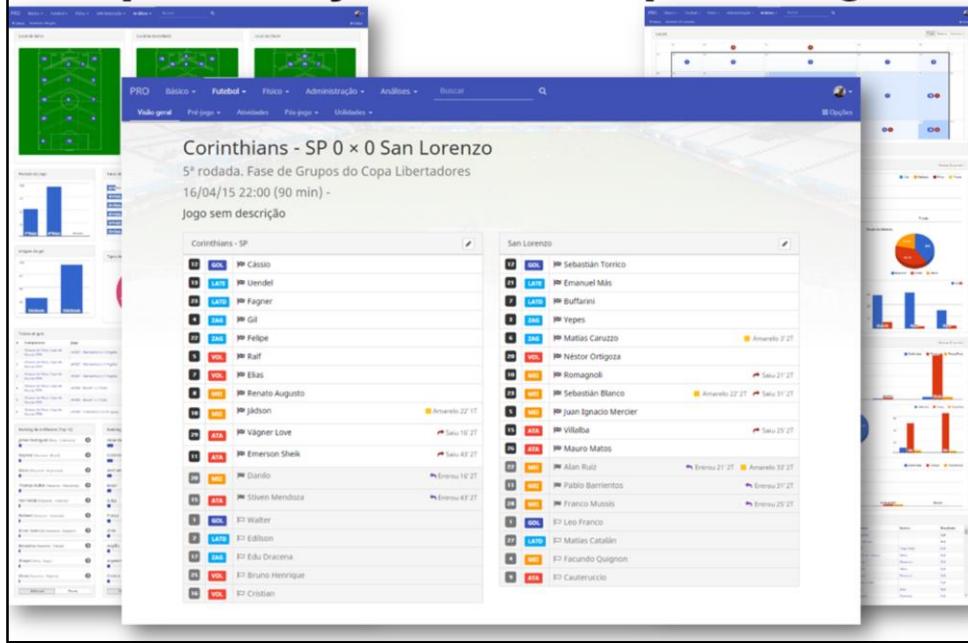
Aqui um exemplo de relatório interativo de gols, com informações sobre cada particularidade dos gols

Apresentação: Antes e após o Jogo



Esse é um relatório dos pênaltis, com detalhes do comportamento do batedor e do goleiro

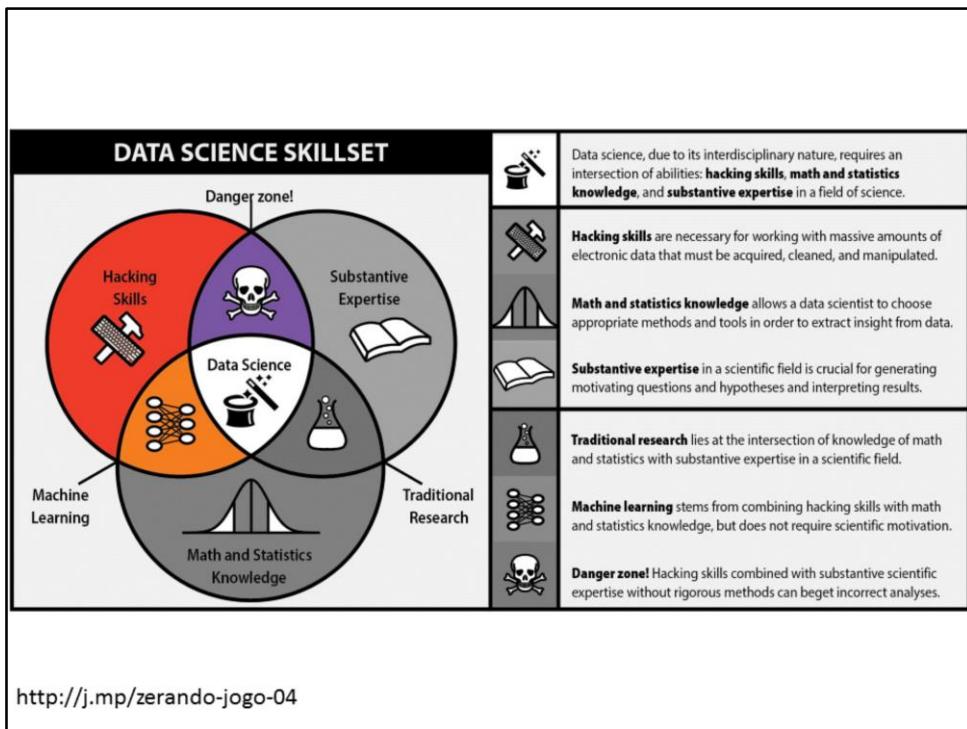
Apresentação: Antes e depois do Jogo



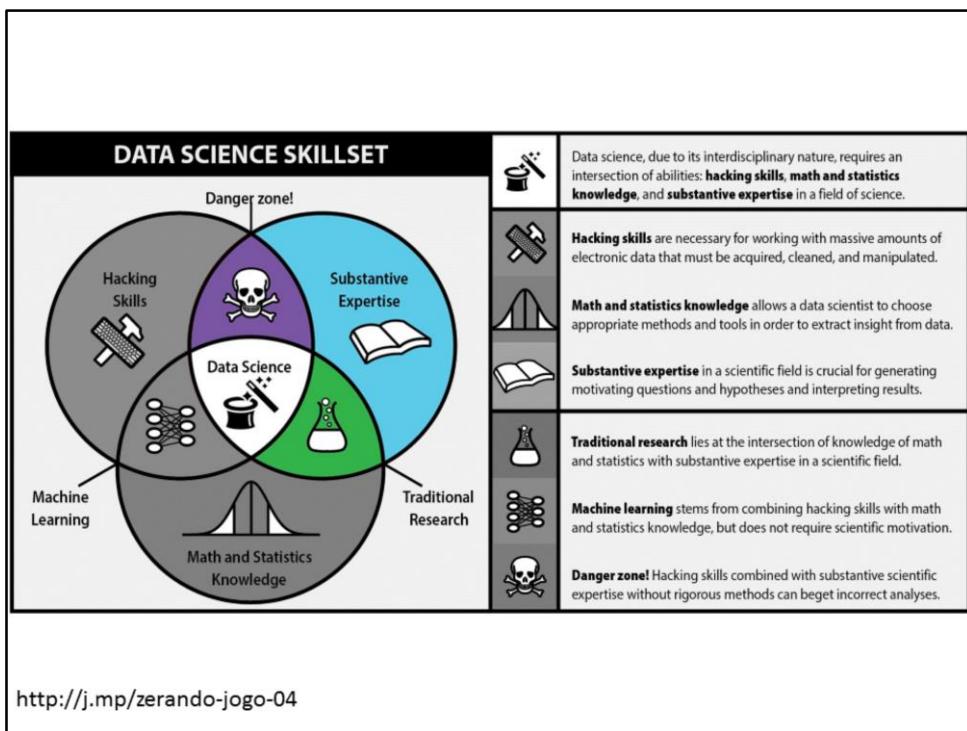
E essa é uma tela de visão do jogo, que mostra todos os lances do jogo com detalhes e observações



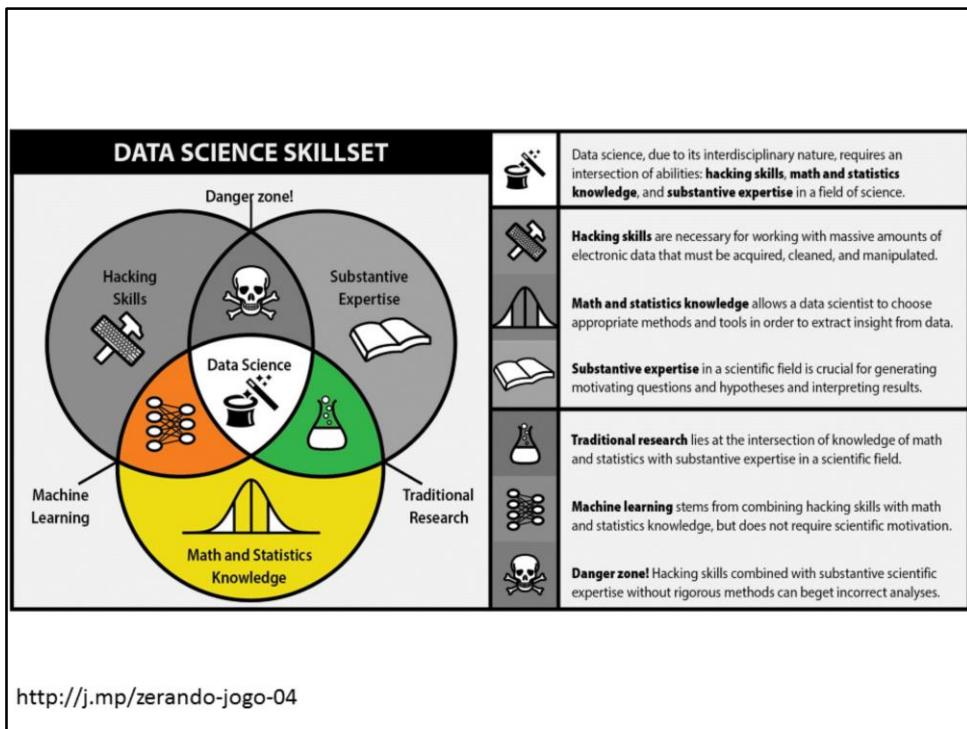
No Corinthians, o CIFUT não é oficialmente uma equipe de Data Science, embora tenha esse papel. Agora vou mostrar o que é necessário em uma equipe de Data Science completa.



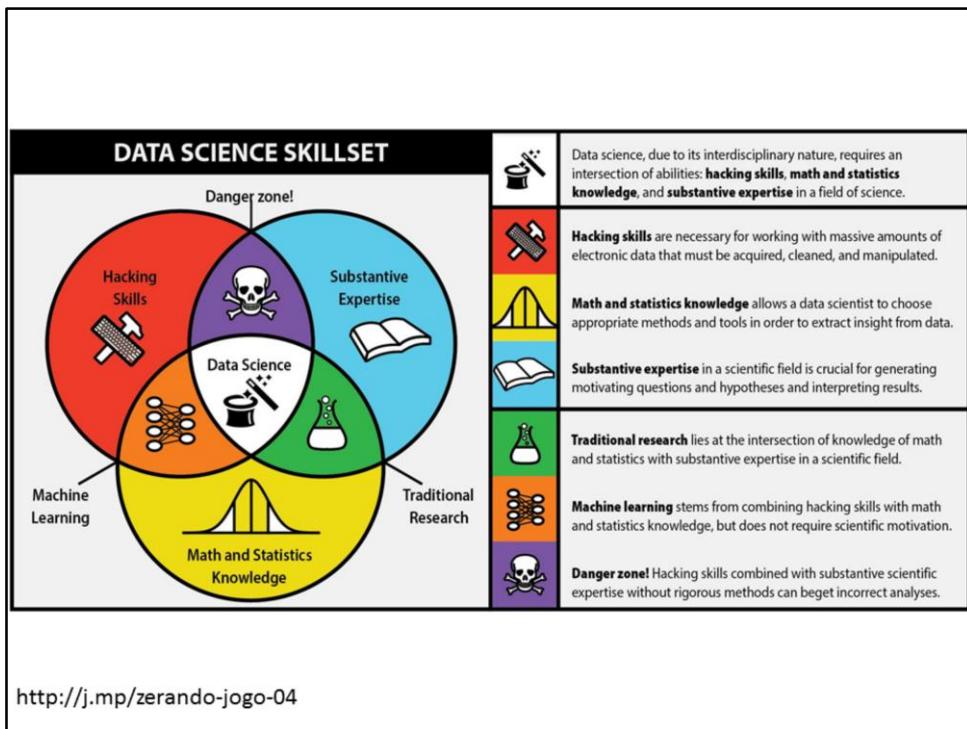
Primeiro, o motivo que nos trouxe aqui hoje, habilidade na ciência da computação, o que dá poder à Data Science é o uso inteligente dos recursos computacionais existentes, precisamos de alguém capaz de transformar as hipóteses em modelos reproduzíveis e que consiga desenvolver formas de manipular grandes volumes de dados.



Precisamos também de conhecimento específico na área de estudo, para conseguir formular as hipóteses e interpretar os resultados da melhor forma possível, a qualidade da informação produzida depende da qualidade do conhecimento da equipe na área estudada.



Por último, conhecimento matemático e estatístico. Conhecer o ambiente e saber usar a tecnologia a favor não é suficiente, é preciso saber como extrair conhecimento dos dados usando matemática, ela é o elo que une a computação com o conhecimento específico.



As três habilidades são muito importantes para formar a Data Science.
 Usar o poder computacional apenas aplicando o conhecimento específico pode produzir conhecimento incorreto, a ciência tradicional combina apenas o conhecimento específico e a matemática, o Machine Learning não precisa necessariamente de um ramo científico para funcionar, só precisa de matemática e poder computacional.



Quem se interessou por esse ramo, recomendo que se aprofunde, isso tende a ser uma área muito concorrida no futuro breve.

Estatística

Vou mostrar agora algumas formas para nós desenvolvedores nos aprofundarmos em algumas áreas da Data Science, primeiro em estatística.



<http://j.mp/zerando-jogo-05>

Esse livro da O'Reilly dá uma visão geral de estatística e análise exploratória de dados usando python.

Estatística

O'REILLY® 2nd Edition

Think Stats

EXPLORATORY DATA ANALYSIS



Allen B. Downey

<http://j.mp/zerando-jogo-05>

UNIVERSITY OF TORONTO + coursera

Statistics, Making sense of Data

<http://j.mp/zerando-jogo-06>

A universidade de Toronto disponibiliza, através do Coursera, uma introdução sobre estatística para programadores. É gratuito e muito prático, recomendo que deem uma olhada no coursera, tem muitos cursos de excelente qualidade gratuitos e com opção de certificação reconhecida.

+

Statistics, Making sense of Data

<http://j.mp/zerando-jogo-06>

<http://j.mp/zerando-jogo-05>

<http://j.mp/zerando-jogo-07>

Também recomendo pesquisar sobre o R, é uma linguagem projetada para estatística, tem excelentes recursos, é bem simples e está sendo considerada como a linguagem nativa da Data Science.

Machine Learning

Machine Learning é uma das áreas de estudo mais fascinantes, pelo menos pra mim
hehe.

Machine Learning

A Programmer's Guide to
Data Mining



The Ancient Art of the Numerati

Ron Zacharski

<http://j.mp/zerando-jogo-08>

Esse guia orienta nós programadores para como praticar a arte de minerar os dados

Machine Learning

A Programmer's Guide to
Data Mining



The Ancient Art of the Numerati

Ron Zacharski

Stanford
University

+ coursera

Machine Learning

<http://j.mp/zerando-jogo-09>

<http://j.mp/zerando-jogo-08>

No coursera existe um curso muito bom da universidade de Stanford, que agora pode ser feito sob demanda a qualquer momento, não precisa esperar uma abertura de turma, também gratuito.

Machine Learning

A Programmer's Guide to
Data Mining



The Ancient Art of the Numerati

Ron Zacharski

<http://j.mp/zerando-jogo-08>



+ coursera

Machine Learning

<http://j.mp/zerando-jogo-09>



<http://j.mp/zerando-jogo-10>

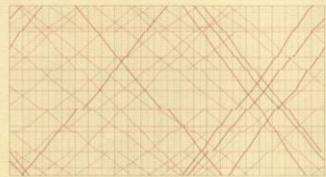
E claro que não poderia deixar de falar no Python, minha linguagem de programação favorita, ela já tem muitos recursos para machine learning, e talvez tenham visto na palestra do Gabriel um pouco do poder dela.

Nos fins de semana escrevo algumas linhas de python pra relaxar hehe.

Visualização

Temos designers aqui? Eu acredito que todos programadores, matemáticos e humanos deveriam ter pelo menos um conhecimento básico nos conceitos de design, isso se prova na hora gerar uma visualização para os seus resultados.

Visualização



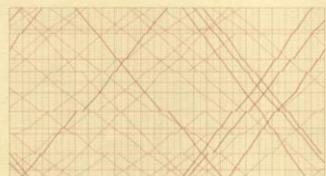
The Visual Display
of Quantitative Information

EDWARD R. TUFTE

<http://j.mp/zerando-jogo-11>

Esse livro do Edward Tufte ensina formas eficientes de apresentar visualmente dados quantitativos.

Visualização



The Visual Display
of Quantitative Information

EDWARD R. TUFTE



RICE®

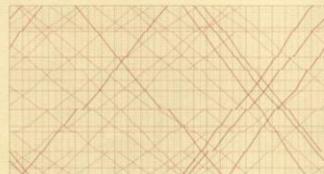
Data Visualization (stat645)

<http://j.mp/zerando-jogo-12>

<http://j.mp/zerando-jogo-11>

Também existe um curso disponível na internet na Universidade do Rice

Visualização



The Visual Display
of Quantitative Information

EDWARD R. TUFTE

<http://j.mp/zerando-jogo-11>



RICE®

Data Visualization (stat645)

<http://j.mp/zerando-jogo-12>



<http://j.mp/zerando-jogo-13>

E pra aproveitar o poder do HTML5 para produzir visualizações recomendo que deem uma olhada no D3, ele é feito originalmente em javascript mas tem port pra Python também

Computação em Escala

A computação em nuvem permitiu que um pequeno investimento fosse suficiente para construir grandes redes computacionais.



E na crista da onda vem surfando o elefante amarelo do hadoop, esse livro é muito bom e deveria ficar na cabeceira da cama de quem quer trabalhar com Big Data

Computação em Escala

Storage and Analysis at Internet Scale

3rd Edition
Revised & Updated

Hadoop

The Definitive Guide

O'REILLY®

Tom White

<http://j.mp/zerando-jogo-14>

ISB + coursera

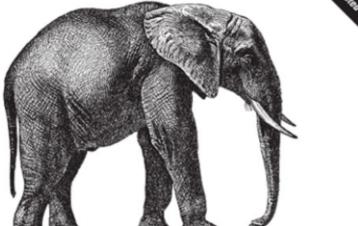
Web Intelligence and Big Data

<http://j.mp/zerando-jogo-15>

No coursera também tem um curso sobre inteligência web e big data

Computação em Escala

Storage and Analysis at Internet Scale
3rd Edition Revised & Updated



Hadoop
The Definitive Guide

O'REILLY* Tom White

<http://j.mp/zerando-jogo-14>

ISB + coursera

Web Intelligence and Big Data

<http://j.mp/zerando-jogo-15>



<http://j.mp/zerando-jogo-16>

Deem uma olhada também no Spark, da fundação Apache, ele é como um hadoop com esteroides.



Pra gente encerrar, é possível zerar o futebol? Encontrar uma estratégia perfeita que vai garantir a vitória em todos os jogos?



Sim, é possível! Mas é muito difícil.

Não foi o caso da seleção da Austrália que fez 31 a 0 contra a Samoa Americana nas eliminatórias da copa de 2002, isso foi uma combinação da habilidade da seleção australiana, da falta de habilidade da Samoa e do acaso que deu uma vantagem insuperável para os vencedores da partida.

Obstáculos para zerar o jogo

Existem obstáculos gigantescos que ainda nos impedem de zerar o jogo.

Obstáculos para zerar o jogo

Como saber se temos dados suficientes?

Não temos como saber se os dados coletados são suficientes para nos permitir formar uma hipótese boa o suficiente.

Obstáculos para zerar o jogo

**Como saber se a pergunta
feita vai dar a resposta que
precisamos?**

E com as hipóteses que conseguirmos formar não podemos garantir também que o resultado produzido vai nos dar algum retorno real

Complexidade do Futebol

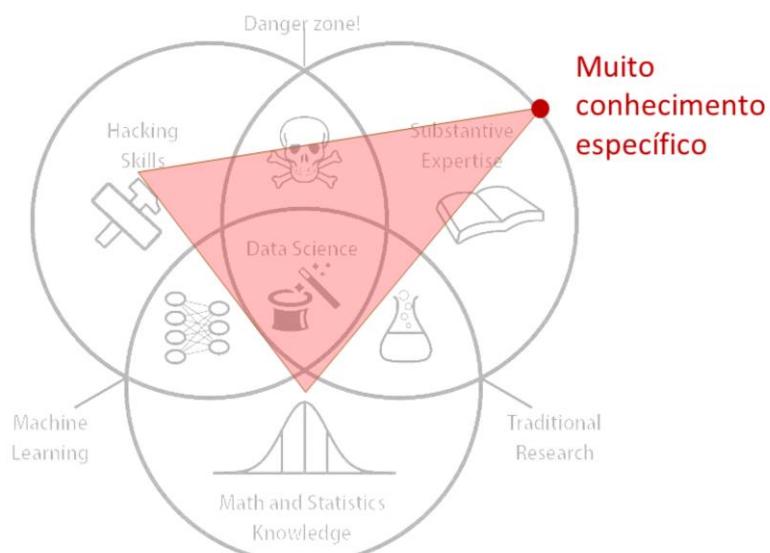
Não sei o que aconteceu ou se aconteceu não tô sabendo.

Mas futebol é isso que aconteceu, porque eu não sei o que aconteceu.

<http://j.mp/zerando-jogo-18>

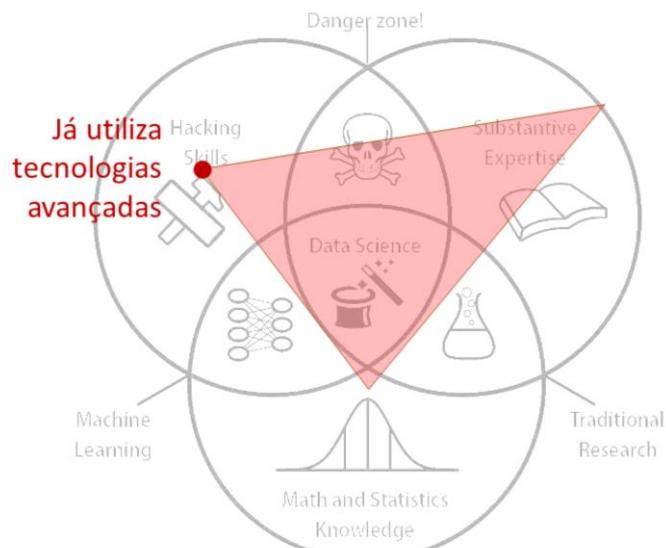
Tudo isso porque o futebol é extremamente complexo e mutável, muito difícil de controlar e com uma cultura que descarta o uso da matemática aplicada.

Data Science no Futebol Hoje



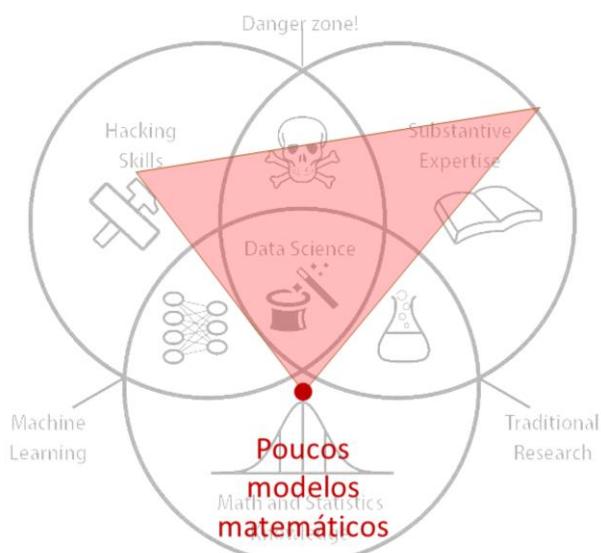
O cenário atual da Data Science no futebol tem um desequilíbrio bem nítido. Por um lado temos muito conhecimento específico nas comissões técnicas dos clubes no mundo inteiro.

Data Science no Futebol Hoje



Existem várias empresas lucrando por usar tecnologias avançadas para entregar informações sobre futebol, embora sejam informações básicas.

Data Science no Futebol Hoje



O elo fraco está na inexistência de modelos matemáticos de qualidade, os modelos que existem são muito básicos, são somas e subtrações se comparados com a evolução dos modelos em outras modalidades esportivas e ramos da ciência. Isso devido principalmente a grande variedade de tipos de evento dentro de um jogo e na escassez desses eventos para analisar e testar.

Solução:



Existe uma solução para nos ajudar a zerar o jogo?

Solução:

Usar a tecnologia para produzir novos modelos matemáticos e testá-los.



Sim, podemos usar o poder da tecnologia para simular os eventos raros e produzir modelos matemáticos mais confiáveis.



Perguntas?



Perguntas?

Sobre a apresentação

