

# Criando produtos de Data Science & AI

Da proposta ao deploy



@wespatrocinio

Twitter | LinkedIn | Facebook | GMail | Skype | SlideShare



## 1. FORMAÇÃO



- Bacharelado em Física Computacional
- Mestrado em Física Aplicada (simulações e HPC);

## 2. PESQUISADOR



**Wernher von Braun**  
centro de pesquisas avançadas

- Pesquisador de novos materiais para microeletrônica e spintrônica



**Wernher von Braun**  
centro de pesquisas avançadas

- Desenvolvedor de software;
- Líder de desenvolvimento de apps mobile e web;
- Líder de desenvolvimento de soluções para IoT;



*Telefonica*

- Data Scientist;
- Gerente de Produtos de Data Science & AI;
- Data Architect;



**Daitan**  
GROUP  
*Accelerating*

## 3. SOFTWARE

## 4. DATA SCIENCE & AI

Mais uma definição...

---

# O que é um *data product*?



# DATA PRODUCTS

*Um data product é um produto que usa dados / informações para facilitar o fornecimento do valor esperado.*

# Roadmap de um *data product*





Não coloque o carro na frente dos bois

---

# Tenha certeza de que está pronto






# Como começar?



- **Alguém precisa do seu produto?**
  - Pergunte isso antes de qualquer esforço e fórmulas matemáticas mirabolantes;
- **Comece** de maneira **simples** o suficiente para que você saiba quem são seus clientes e se você pode resolver suas necessidades;
  - *Minimum viable product*;
- **Mantenha simples** o maior tempo possível;
  - Do simples para o complexo;



A person wearing a white jumpsuit with a red sash and black sneakers stands next to a vintage car. A sign on the car reads 'RACIONAIS'. The background is a blue sky with clouds.

# Nada como um dia após o outro dia

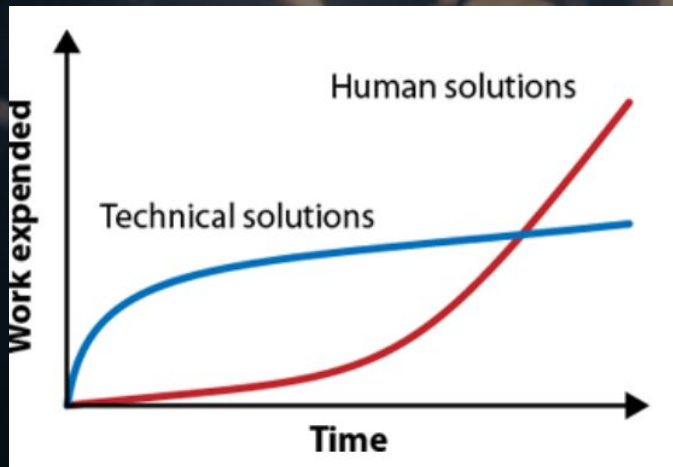
- Tenha certeza que o desenvolvimento é **viável**
  - Desbloqueie governança, PII, segurança, etc);
- Um *timing* errado pode deixar o **time desmotivado** ou gerar **expectativa desalinhada** no cliente;



# Defina os papéis

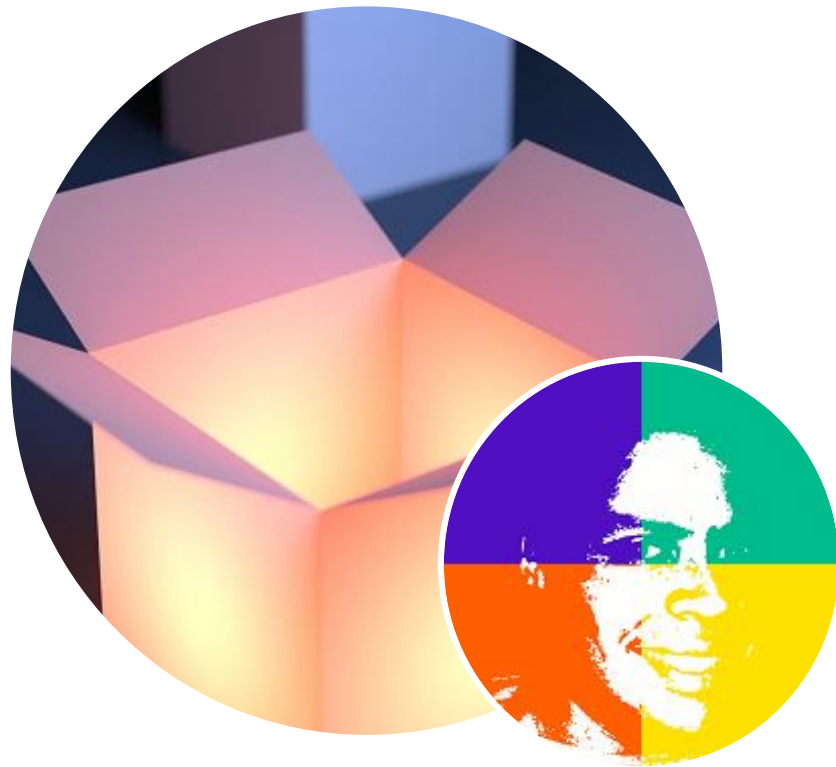
- Um *data product* possui **várias camadas de escopo completamente diferentes** (infra, APIs, bancos de dados, etc);
- Defina e envolva todos os *players* que terão responsabilidade no desenvolvimento **desde o princípio**;
  - Não estou dizendo que deve ser "*Corp-wide*"
  - *Squads*;
- Não caia no conto do *handover* após a publicação da primeira versão;

# Use pessoas para começar



- Use pessoas para conhecer melhor o seu problema / desafio;
- Soluções técnicas / automáticas sempre serão mais escaláveis a longo prazo, mas ...
- Quando você está começando, o longo prazo não deve ser tão importante;
  - No curto prazo, pessoas exigem menos trabalho;
- Preocupe-se com escala quando você precisar dela;

# Prepare os dados





# Domínio de dados

- Entenda o **significado** das variáveis que estão sendo consideradas, assim como as **regras** que a definem;
- Defina as **propriedades de negócio**;
  - O que é *churn*?
- Nem todo KPI padrão é construído da mesma forma;
  - ARPU, *ticket* médio, *customer lifecycle value*, etc;
- Metadados



# Democratização

- Dados **acessíveis & vigentes** durante o tempo necessário;
- Tratamento de **qualidade** dos dados;
- Mínimo de **padronização & estabilidade** nos dados que servirão de insumo;




# Métricas e metas





Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.  
Metricas e metas do produto antes do desenvolvimento.



A yellow measuring tape is laid diagonally across a wooden surface. The tape has black markings and numbers. The background is slightly blurred, emphasizing the tape and the text overlaid on it.

# Métricas e metas

Estamos falando de **NEGÓCIOS**.

Serão o seu principal **guia** para os  
esforços do desenvolvimento!

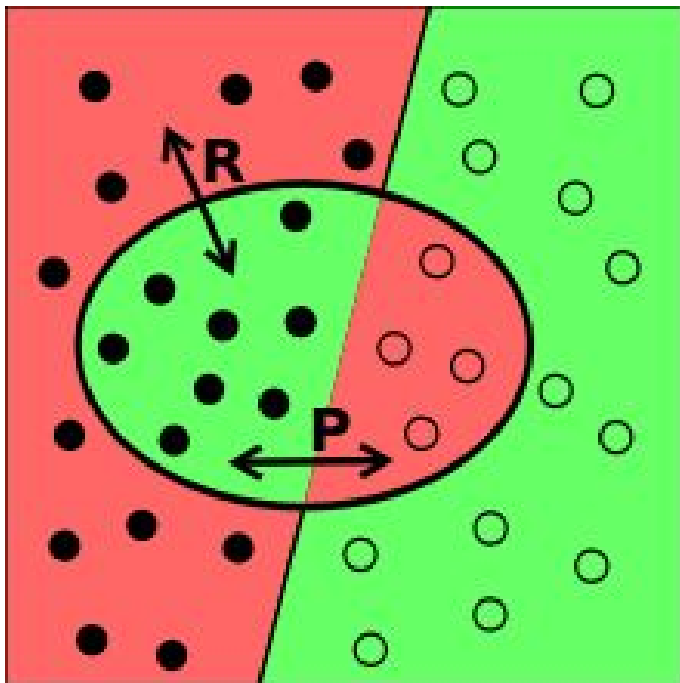
Primeiro gere o valor

---

# Desenvolva a inteligência



# Precisão VS *recall*



- **Precisão:** exatidão em prover o que é desejado;
- **Recall:** conjunto de boas classificações (inventário);
- Princípio de incerteza!
  - Você sai da Física, mas a Física nunca sai de você;
- Quando há expectativa, precisão é fundamental;
  - Usuário ativo
  - Por exemplo, buscas;
- Em ads, nem tanto;
  - Não há expectativa;



# Avaliação offline

- Crie uma emulação de seu cenário real de otimização com dados passados;
  - Quantidade de tempo / dados relevantes;
- Tenha certeza de que o cenário não terá mudança significativa no curto / médio prazo
  - Mudança de portfólio de produtos, mudança no modelo de negócio, etc;
- Avaliações offline tendem a subestimar o resultado de um teste A/B;



# Performance

- Busque apenas o ganho de inteligência;
  - Tenha em mente a arquitetura e SLAs, mas mantenha o foco na inteligência;
- Teste o tempo de vida de sua inteligência para determinar o seu ciclo de vida;
- Teste a performance em subdomínios específicos para checar a necessidade de fragmentar a inteligência;

# ***Minimum Viable Product***







# Desacoplado

- Desacople ao máximo a inteligência de sua interface de consumo
  - API e modelo devem ser independentes e exigem atualizações em momentos / situações diferentes;
- REST
  - *Stateless*;
  - Cache / tempo real;
  - Separar responsabilidades *client* / *server*;

A photograph of the Space Shuttle Columbia being launched from the launch pad. The shuttle is white with black and red markings, and it is surrounded by a large plume of white smoke and fire. The launch pad structure is visible on the left side of the image.

# Escalável

- A lenda do protótipo que vira produto;
  - Não precisa se preocupar com detalhes de monitoramento, notificações, etc. Mas seja rapidamente escalável;
- Não subestime o poder de uma AI bem feita;
  - Engajamento de usuários & eventos;
  - *Uplifts* de 300%;

# Teste A/B





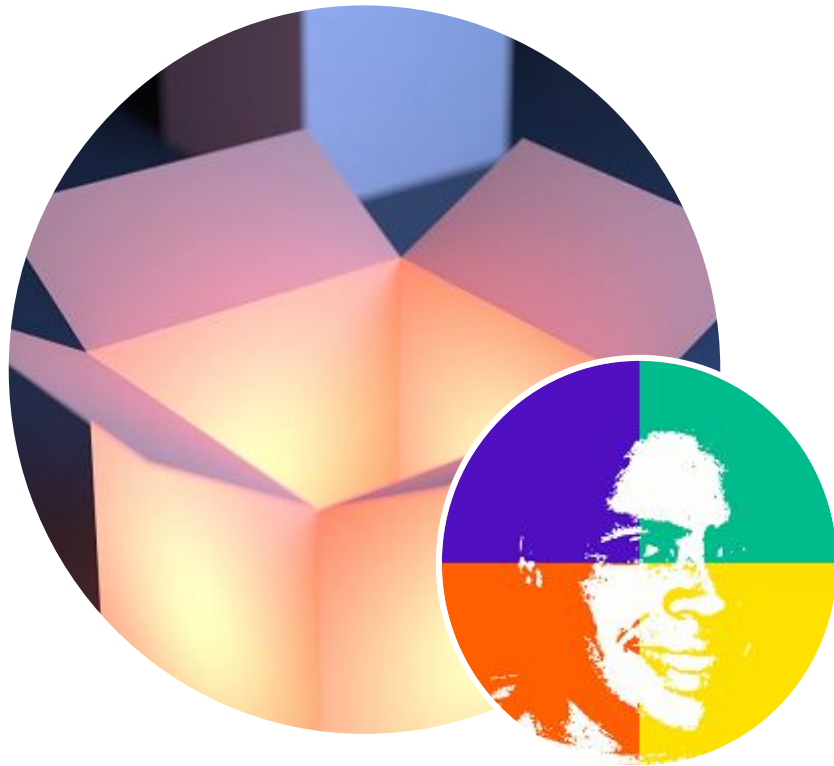
# Experimento

- Método científico: uma hipótese de cada vez;
- É A | B, e não ABCDEFGHJDH;
- Teste estatisticamente relevante
  - Duração, quantidade de usuários, nichos, etc;
- Foco nas KPIs do teste, mas mantenha os olhos em outros indicadores importantes do produto;
  - Evitar o "cobertor pequeno";

Brace yourselves

---

# Catch-up







Vai que...

A man with a beard and long hair, wearing a dark fur cloak, is shown in profile, looking down and holding a sword. The text "Catch up" is overlaid in a large, white, italicized font.

# *Catch up*

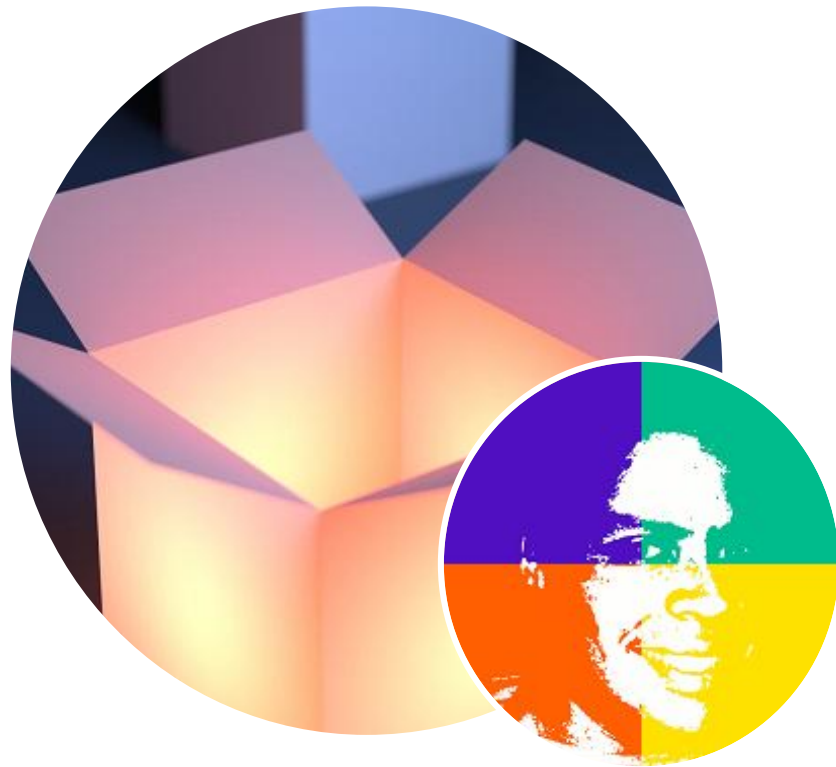
- Avaliações offline tendem a subestimar o resultado de um teste A/B, lembra?
- Volte para a avaliação offline e aumente a performance da inteligência;
- Seja ágil (se você fez um bom desacoplamento, isso será fácil);
- Mais dados *versus* diferentes algoritmos;



Agora a p\* ficou séria

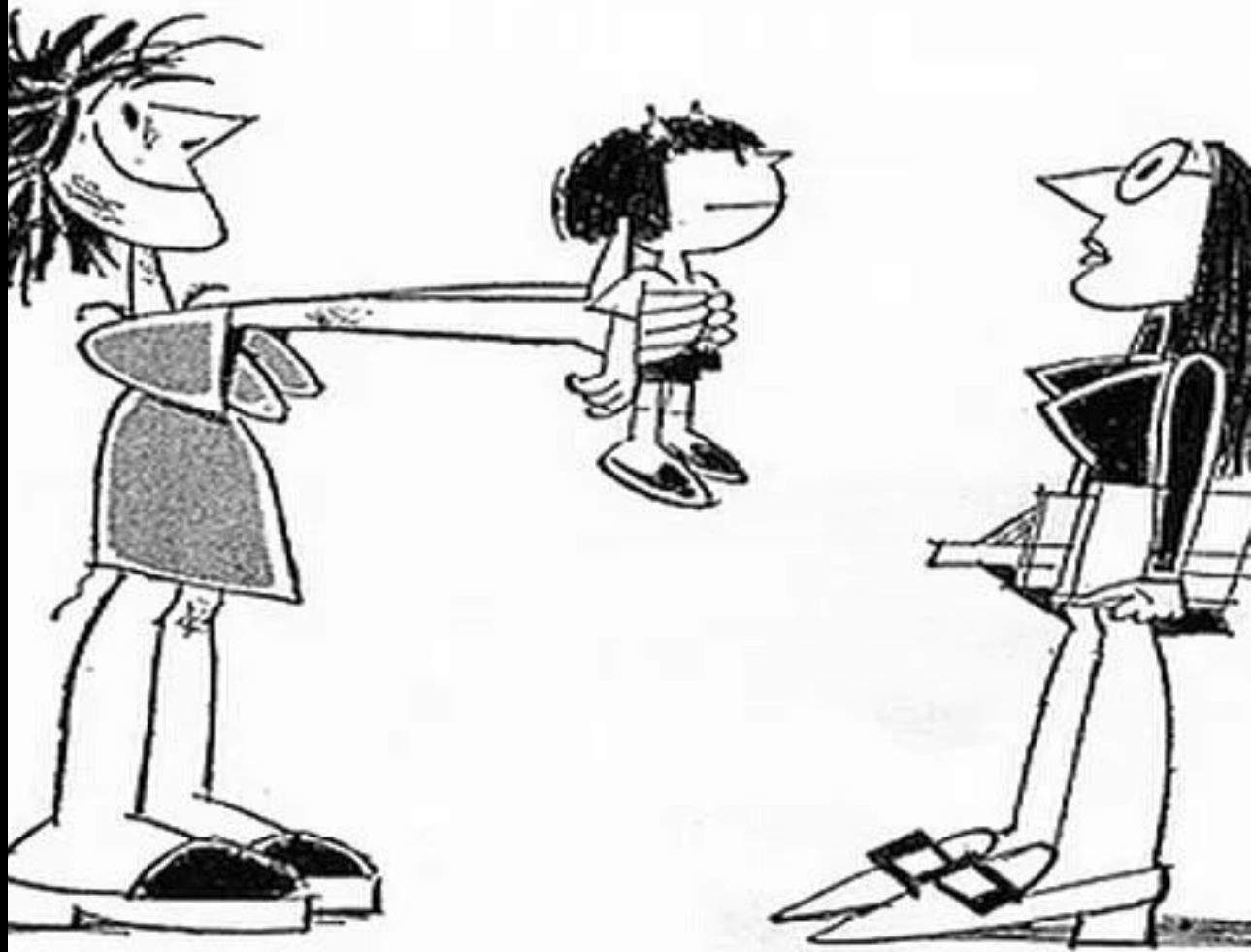
---

# Produção



## Produção

Espero que você tenha definido e mantido bem os que deveriam estar envolvidos.





# Operação

- Monitoramento;
  - DevOps & BizOps;
- Alertas;
- Automação de deploy;
- Utilize a operação real para melhorar ainda mais sua inteligência;



*That's all Folks!*



# OBRIGADO!

Perguntas?

**@wespatrocinio**

Twitter | LinkedIn | Facebook | GMail | Skype | SlideShare