

Implantação & Monitoramento de Sistemas de ML

Kunumi - 2020



Nosso propósito é
desenvolver Inteligência
Artificial com responsabilidade
para distribuir conhecimento,
prosperidade e bem-estar.

SOBRE MIM

Lucas (you) ● Lucas de Miranda Bastos



2010

UF *m* G

2013



UF *m* G

2017



SOBRE MIM

Lucas (you) ● Lucas de Miranda Bastos



2010



UF m G

2013



KUNUMI
UF m G

2017



||<< MOTIVAÇÃO

“Em um belo dia ensolarado, você e sua equipe estão prontos para implantar o sistema que acabaram de produzir no ambiente do cliente. As métricas coletadas dos experimentos utilizando o sistema são incríveis e vocês tem certeza que o sistema trará grandes impactos para as pessoas.”

||<< MOTIVAÇÃO

“Em um belo dia ensolarado, você e sua equipe estão prontos para implantar o sistema que acabaram de produzir no ambiente do cliente. As métricas coletadas dos experimentos utilizando o sistema são incríveis e vocês tem certeza que o sistema trará grandes impactos para as pessoas.”

- **Dia 0:** Você tem problemas para fazer o sistema funcionar em produção. Mas isso é estranho, afinal “ele funciona na minha máquina!”™.

||<< MOTIVAÇÃO

“Em um belo dia ensolarado, você e sua equipe estão prontos para implantar o sistema que acabaram de produzir no ambiente do cliente. As métricas coletadas dos experimentos utilizando o sistema são incríveis e vocês tem certeza que o sistema trará grandes impactos para as pessoas.”

- **Dia 0:** Você tem problemas para fazer o sistema funcionar em produção. Mas isso é estranho, afinal “ele funciona na minha máquina!”™.
- **Dia 1:** Depois de um dia corrido, você recebe uma mensagem falando que o sistema parou de funcionar no meio da madrugada. E agora? Novamente: “ele funciona na minha máquina!”™

||<< MOTIVAÇÃO

“Em um belo dia ensolarado, você e sua equipe estão prontos para implantar o sistema que acabaram de produzir no ambiente do cliente. As métricas coletadas dos experimentos utilizando o sistema são incríveis e vocês tem certeza que o sistema trará grandes impactos para as pessoas.”

- **Dia 0:** Você tem problemas para fazer o sistema funcionar em produção. Mas isso é estranho, afinal “ele funciona na minha máquina!”™.
- **Dia 1:** Depois de um dia corrido, você recebe uma mensagem falando que o sistema parou de funcionar no meio da madrugada. E agora? Novamente: “ele funciona na minha máquina!”™
- **Dia 2:** Você descobre que o código tinha um *bug* e precisa enviar uma correção, porém, percebe que alguns dos problemas de incompatibilidade do dia 0 voltaram a aparecer... Maaaaas... “ele funciona na minha máquina!”™

||<< MOTIVAÇÃO

“Em um belo dia ensolarado, você e sua equipe estão prontos para implantar o sistema que acabaram de produzir no ambiente do cliente. As métricas coletadas dos experimentos utilizando o sistema são incríveis e vocês tem certeza que o sistema trará grandes impactos para as pessoas.”

- **Dia 0:** Você tem problemas para fazer o sistema funcionar em produção. Mas isso é estranho, afinal “ele funciona na minha máquina!”™.
- **Dia 1:** Depois de um dia corrido, você recebe uma mensagem falando que o sistema parou de funcionar no meio da madrugada. E agora? Novamente: “ele funciona na minha máquina!”™
- **Dia 2:** Você descobre que o código tinha um *bug* e precisa enviar uma correção, porém, percebe que alguns dos problemas de incompatibilidade do dia 0 voltaram a aparecer... Maaaaas... “ele funciona na minha máquina!”™
- **Dia 60:** Você recebe um email falando que o sistema já não é mais o mesmo do dia em que foi colocado em produção. A performance caiu, o impacto é menor, mas nada mudou... Ou mudou?

« MOTIVAÇÃO

“Em um belo dia ensolarado, você e sua equipe estão prontos para implantar o sistema que acabaram de produzir no ambiente do cliente. As métricas coletadas dos experimentos utilizando o sistema são incríveis e vocês tem certeza que o sistema trará grandes impactos para as pessoas.”

- **Dia 0:** Você tem problemas para fazer o sistema funcionar em produção. Mas isso é estranho, afinal “ele funciona na minha máquina!”™.
- **Dia 7:** Depois de um dia corrido, você recebe uma mensagem falando que o sistema parou de funcionar no meio da madrugada. Agora? Novamente: “ele funciona na minha máquina!”™
- **Dia 15:** Você descobre que o código tinha um *bug* e precisa enviar uma correção, porém, percebe que alguns dos problemas de incompatibilidade do dia 0 voltaram a aparecer... aaaaaas... “ele funciona na minha máquina!”™
- **Dia 60:** você recebe um email falando que o modelo já não é mais o mesmo do dia em que foi colocado em produção. A performance caiu, o impacto é menor, mas nada mudou... Ou mudou?

||<< **INTRODUÇÃO**

No geral, o conteúdo disponível sobre Aprendizado de Máquina é majoritariamente focado na etapa de ajuste do modelo.

||<< INTRODUÇÃO

No geral, o conteúdo disponível sobre Aprendizado de Máquina é majoritariamente focado na etapa de ajuste do modelo.

- Experimental Design / ML in the Real World

- ▶ Deploying Models to Real-Time Systems ▲
High-level thoughts on various ways to deploy your trained models to production systems including apps and websites.

- ▶ A/B Testing Concepts ▼

- ▶ T-Tests and P-Values ▼

- ▶ [Activity] Hands-on With T-Tests ▼

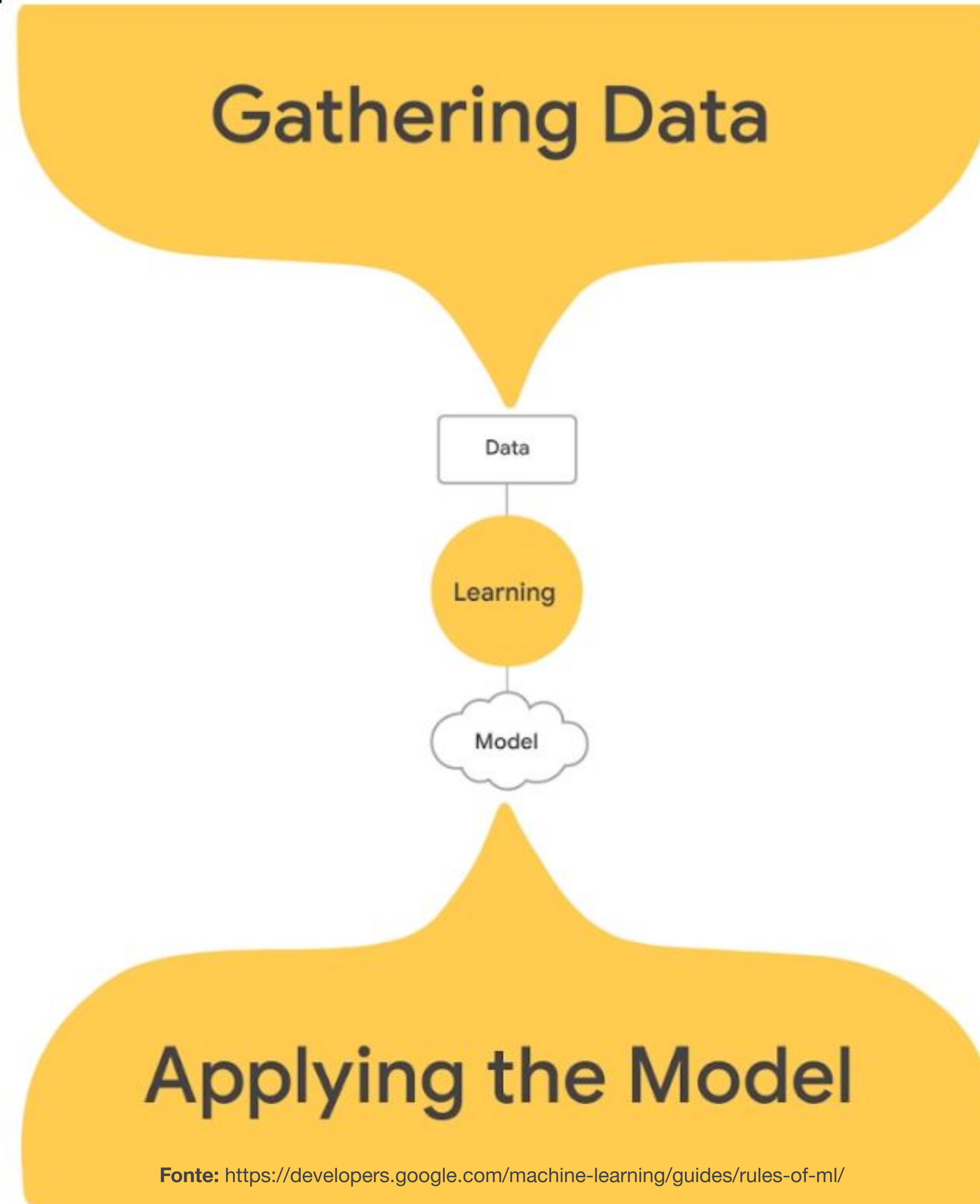
- ▶ Determining How Long to Run an Experiment ▼

- ▶ A/B Test Gotchas ▲
There are many limitations associated with running short-term A/B tests - novelty effects, seasonal effects, and more can lead you to the wrong decisions. We'll discuss the forces that may result in misleading A/B test results so you can watch out for them.

||<< INTRODUÇÃO



Fonte: <https://developers.google.com/machine-learning/guides/rules-of-ml/>



Implantação

◀ IMPLANTAÇÃO

Se o sistema vai ser adicionado a um outro sistema já existente, como eles vão se comunicar?

- Banco de dados
- API
- Fila de requisições

||<< IMPLANTAÇÃO

Se o sistema vai ser adicionado a um outro sistema já existente, como eles vão se comunicar?

- Banco de dados
- API
- Fila de requisições

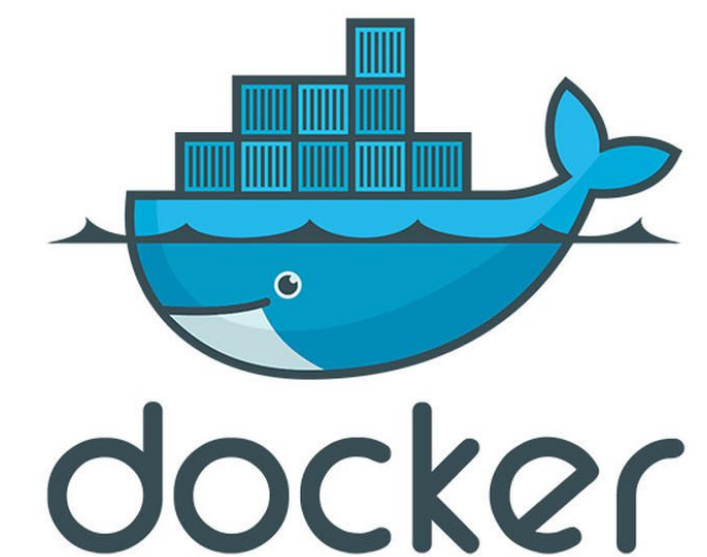
Se o modelo existe por si só, como ele é acionado?

- Desenvolver uma página na web ou aplicativo?
- Desenvolver um software de linha de comando?

◀ IMPLANTAÇÃO

Qual a melhor estratégia para enviar o sistema pro ambiente de produção?

- O arquivo do modelo será armazenado junto com o código? Separado?
 - Se ele está separado, como faço *download* dele?
- Como evitar o problema de “mas roda na minha máquina!”?
 - Utilizar *containers* de imagens, como *Docker*.
 - Ambiente de homologação espelhado no ambiente de produção.



||<< IMPLANTAÇÃO

Qual a melhor estratégia para enviar o sistema pro ambiente de produção?

- O arquivo do modelo será armazenado junto com o código? Separado?
 - Se ele está separado, como faço *download* dele?
- Como evitar o problema de “mas roda na minha máquina!”?
 - Utilizar *containers* de imagens, como *Docker*.
 - Ambiente de homologação espelhado no ambiente de produção.

O sistema vai ser enviado para um ambiente na Nuvem ou um Datacenter próprio?

- Qual o impacto dessa decisão no orçamento?
- Como sincronizar todos os serviços na Nuvem.

||<< IMPLANTAÇÃO

Qual a melhor estratégia para enviar o sistema pro ambiente de produção?

- O arquivo do modelo será armazenado junto com o código? Separado?
 - Se ele está separado, como faço *download* dele?
- Como evitar o problema de “mas roda na minha máquina!”?
 - Utilizar *containers* de imagens, como *Docker*.
 - Ambiente de homologação espelhado no ambiente de produção.

O sistema vai ser enviado para um ambiente na Nuvem ou um Datacenter próprio?

- Qual o impacto dessa decisão no orçamento?
- Como sincronizar todos os serviços na Nuvem.

Nota: Um ambiente de homologação é um ambiente espelhado no ambiente de produção.

◀ IMPLANTAÇÃO

Qual a periodicidade de funcionamento do sistema?

- Diário, Semanal, Mensal, etc...

◀ IMPLANTAÇÃO

Qual a periodicidade de funcionamento do sistema?

- Diário, Semanal, Mensal, etc... **Agendador de Tarefas**
 - Ele se adapta bem aos requisitos **do seu** projeto?
 - Como ele lida com imprevistos?

◀ IMPLANTAÇÃO

Qual a periodicidade de funcionamento do sistema?

- Diário, Semanal, Mensal, etc... **Agendador de Tarefas**
 - Ele se adapta bem aos requisitos **do seu** projeto?
 - Como ele lida com imprevistos?
- *On demand*. Ou seja, uma entidade vai invocar a rotina de execução do sistema.
 - O modelo consegue processar múltiplas instâncias simultaneamente?

Monitoramento

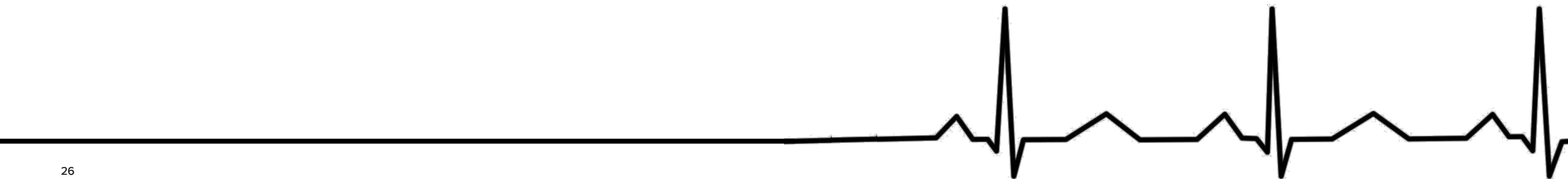
||<< MONITORAMENTO

Primeiro, precisamos saber se um sistema, no ponto de vista de *Software*, está funcionando.

||<< MONITORAMENTO

Primeiro, precisamos saber se um sistema, no ponto de vista de *Software*, está funcionando.

- *Heartbeat*



||<< MONITORAMENTO

Primeiro, precisamos saber se um sistema, no ponto de vista de *Software*, está funcionando.

- *Heartbeat*



||<< MONITORAMENTO

Primeiro, precisamos saber se um sistema, no ponto de vista de *Software*, está funcionando.

- *Heartbeat*
- Notificações

|| MONITORAMENTO

Primeiro, precisamos saber se um sistema, no ponto de vista de *Software*, está funcionando.

- *Heartbeat*
- Notificações

incoming-webhook APP 13:39

Hey, this is your **MINUTELY** notification digest for anomaly_detector.
There are a total of **2** notifications.



Training Error

This notification was triggered 2 times.

Another 0 notifications were triggered in this interval.

||<< MONITORAMENTO

Outra forma de monitoramento é a implementação de *logs* em pontos cruciais do sistema.

O *log* é muito mais poderoso que as funções de `print`.

- Múltiplos “destinos”. (Terminal, arquivos, banco de dados).
- Prioridade de erros (**Debug**, **Informação**, **Aviso**, **Erro**, **Fatal**)

||<< **MONITORAMENTO**

Se o sistema está operando sem problemas, é hora de monitorar sua eficiência.

||<< MONITORAMENTO

Se o sistema está operando sem problemas, é hora de monitorar sua eficiência.

Quais métricas utilizar?

||<< MONITORAMENTO

Se o sistema está operando sem problemas, é hora de monitorar sua eficiência.

Quais métricas utilizar?

Se o sistema tem um objetivo, uma boa ideia é começar investigando-o.

||<< MONITORAMENTO

Se o sistema está operando sem problemas, é hora de monitorar sua eficiência.

Quais métricas utilizar?

Se o sistema tem um objetivo, uma boa ideia é começar investigando-o.

E claro, investigar as métricas clássicas.

||<< MONITORAMENTO

Se o sistema está operando sem problemas, é hora de monitorar sua eficiência.

Quais métricas utilizar?

Se o sistema tem um objetivo, uma boa ideia é começar investigando-a.

E claro, investigar as métricas clássicas.

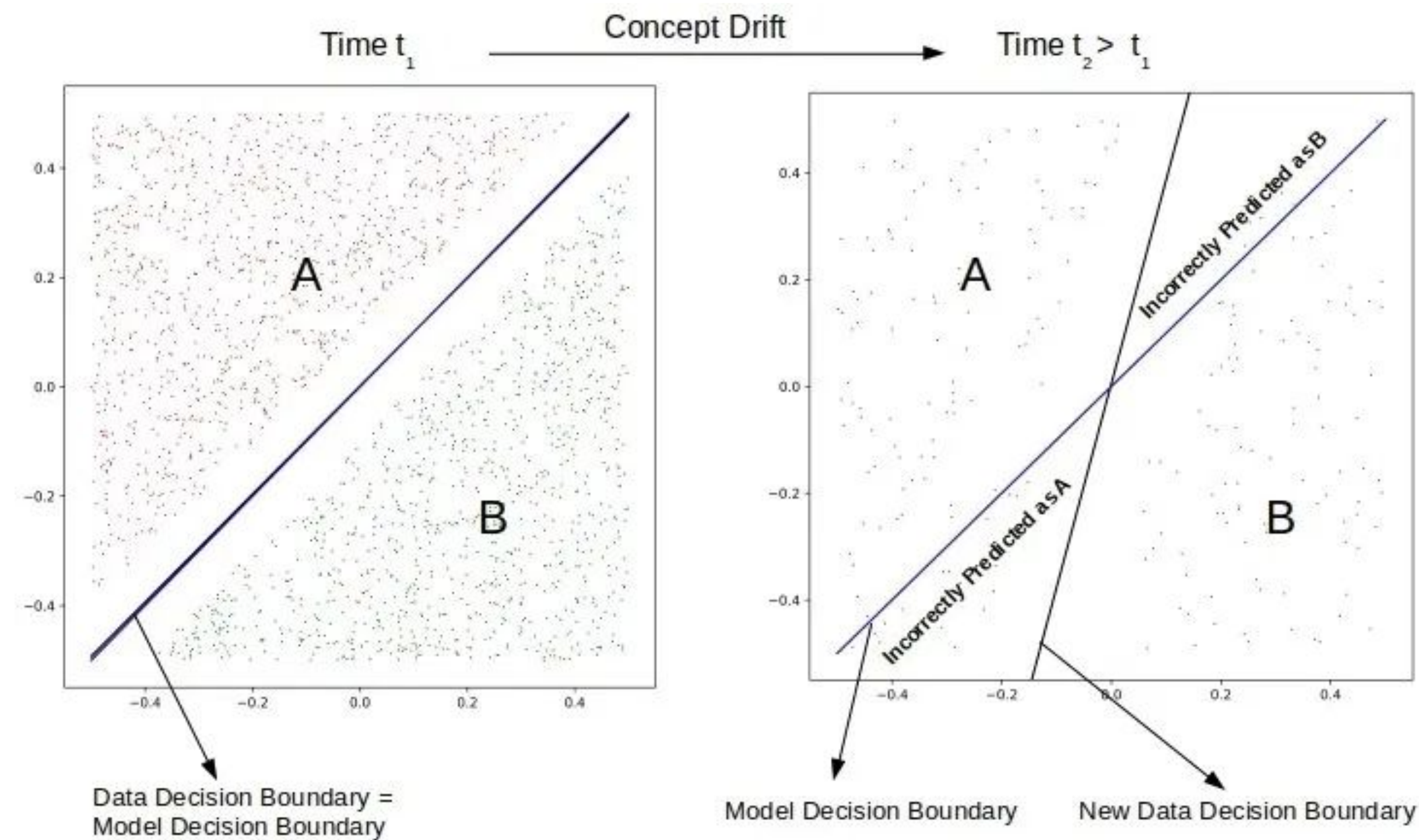
Note que nesse ponto, já temos dados de **produção**. Então estamos lidando com o mundo real.

||<< MONITORAMENTO

Monitorar a eficiência do sistema é interessante para saber se o mesmo está degradando.

MONITORAMENTO

Monitorar a eficiência do sistema é interessante para saber se o mesmo está degradando.



||<< MONITORAMENTO

Monitorar a eficiência do sistema é interessante para saber se o mesmo está degradando.

Quais ações devem ser tomadas quando vê-se o sistema degradando?

- Re-ajustar o sistema?
 - De quanto em quanto tempo?
 - Os novos dados agora podem estar enviesados pela atuação do modelo.
 - Adicionar novas *features*?

MONITORAMENTO



||<< CONCLUSÕES

Desenvolver um bom sistema de ML não é a última etapa de um projeto de Ciência de Dados. O esforço que vem depois da implantação é tão importante quanto o que vem antes.

Coletar informações do desempenho do sistema é fundamental para manutenções futuras e principalmente para não tomar nenhum susto.

KUNUMI

Evolving Intelligence

Obrigado!
lucas@kunumi.com