



Como utilizar um modelo

Quando se cogita a criação de um modelo é importantíssimo saber para quê e como este modelo será utilizado. Este e-book detalha as 3 razões comuns da utilização de um modelo de machine learning.

Data: janeiro/2019

Divirta-se aprendendo de verdade!

www.asn.rocks

Divirta-se aprendendo de verdade!



Adriana Silva

Sou uma eterna apaixonada por ensinar e compartilhar o pouco que sei. Amo estatística e tenho como objetivo de vida mostrar para o mundo o quanto isso é sexy, incrível e esclarecedor. Tenho certeza que a ciência irá mudar nossas vidas! Quer me ajudar nessa missão?



[/in/adrianamms](https://www.linkedin.com/in/adrianamms)



dri@asn.rocks



www.asn.rocks



SUMÁRIO

- 01** INTRODUÇÃO
- 02** AS 3 UTILIDADES DE UM MODELO
- 03** APLICAÇÃO DE NEGÓCIO
- 04** PAPEL DO DATA SCIENTIST
- 05** CONCLUSÃO





01.

INTRODUÇÃO



INTRODUÇÃO

Atualmente existe uma procura frenética por profissionais que sejam capazes de analisar dados e tirar informações valiosas e competitivas dos inúmeros conjuntos de dados disponíveis.

Não é por menos que existe tanta procura por conteúdos que ensinem e detalhem coisas técnicas de como ser um bom Data Scientist. Porém o mais comum que vejo na internet são inúmeros cursos de ferramentas, algumas técnicas, mas ainda sinto falta de conteúdos mais aplicados, realistas, que mostrem estudos reais e que contêm o real trabalho de um Cientista de Dados.

Este e-book tem o objetivo de discutir uma visão mega importante para o cientista de dados, que é o entendimento do problema em que está inserido. E, principalmente, o propósito de seu trabalho!

Exatamente isso! Vamos discutir sobre as possibilidade de aplicabilidade de modelos e detalhar a diferença entre elas, para que você possa sentir o real valor do trabalho de um data scientist e, conseqüentemente, agregar valor aos seus estudos e, principalmente, na sua vida real.

Todo modelo nasce para ser utilizado e, então, porque não se informar como e para quê este modelo será utilizado?





02.

AS 3 UTILIDADES DE UM MODELO



AS 3 UTILIDADES DE UM MODELO



Todo Cientista de Dados está em busca de solucionar algum problema. O único problema que vejo no mercado de trabalho hoje é o fato do “nerd” estar longe da aplicabilidade do que ele irá gerar.

Isso traz uma série de consequências tristes e que, na minha visão, enfraquecem a ciência de dados como um todo.

Como é possível construir algo para resolver o problema de alguém, se eu não sinto “na pele” este problema? Vamos tentar detalhar mais... Imagine que você foi contratado para fazer um modelo, que irá orientar o pessoal de marketing a selecionar o público para uma ação específica.

O Cientista de Dados, muito habilidoso, recebe um conjunto de dados e começa a trabalhar. Desenvolve um algoritmo e entrega o resultado. O pessoal de marketing fica impressionado com a mágica que o cientista de dados fez e está louco para testar tudo.

Aparentemente estão todos felizes e satisfeitos e então o cientista de dados sai para outro desafio. Mas pera aí, onde está o cientista de dados durante a aplicação do algoritmo criado? Ele sabia o real motivo e propósito daquilo? O porquê e para quê aquele algoritmo realmente seria utilizado? Eram claras as limitações de negócio na utilização do algoritmo?

Repare que existem uma série de pontos que devem ser pensados e conversados. Ou seja, o cientista de dados DEVE interagir fortemente com a(s) área(s) de negócio(s) interessada(s).

Vamos ao que importa. Quais são as 3 utilidades de um modelo?

1) DECISÃO

Neste ponto, temos modelos que são capazes de prever um fato com precisão, ou seja, que tem uma taxa de resposta alta. O que isso quer dizer? Quer dizer que o modelo criado será utilizado, sem restrições, e alguma ação será tomada em todos os indivíduos que forem classificados como o evento em questão.

2) ORDENAÇÃO

Já neste caso, temos modelos que são capazes de ordenar a probabilidade de um evento acontecer. O que isso quer dizer? Quer dizer que o modelo criado será utilizado, com algumas restrições de verba por exemplo, e alguma ação será tomada apenas nos indivíduos mais propensos a cometerem o evento em questão. Repare que neste caso a ordem importa, diferente do anterior que se importava com a decisão tomada.

3) ESTIMAÇÃO

No terceiro e último caso, temos modelos que são precisos suficientes para estimar a probabilidade de algo acontecer sem muito erro, ou seja, a estimativa deve estar bem próxima da realidade. O que isso quer dizer? Quer dizer que o modelo criado será utilizado com o objetivo de multiplicar essa probabilidade por algum outro valor, por isso a tamanha importância do menor erro possível.



03.

APLICAÇÃO DE
NEGÓCIO



APLICAÇÃO DE NEGÓCIO

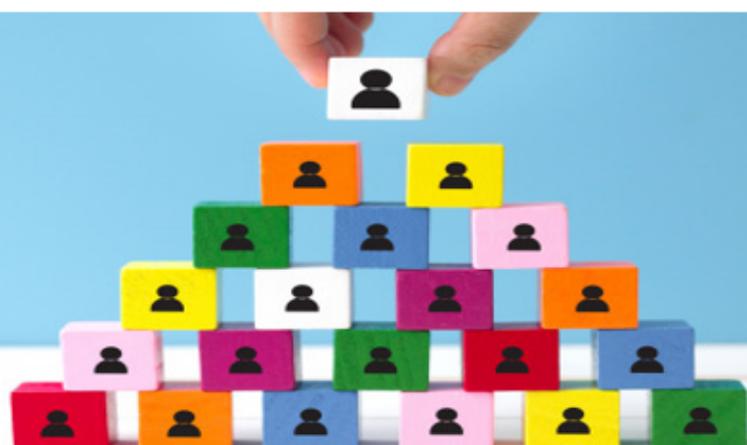
Por mais que as 3 utilidades de um modelo estejam explicadas no capítulo anterior, nada melhor do que ver uma aplicação real para total absorção do tema. Como hoje eu estive em contato com muitas pessoas de RH, estou com isso em mente, e então vou contar aplicações neste cenário.

Sendo assim, convido você a pensar como um diretor de RH, que tem metas de retenção de funcionários, que tem preocupações sérias quando a empresa não está bem financeiramente, e que deve se preocupar com a qualidade do ambiente de trabalho para que não existam processos trabalhistas.

Imagino que ao ler o parágrafo anterior seu cérebro já bombou de ideias de possíveis aplicações de análises para poder atingir nossas metas. E isso está corretíssimo! Pensar analiticamente agrega muito valor!

Porém, vamos contextualizar o primeiro cenário. Nós, diretores de RH, estamos focados na retenção de nossos funcionários. Seria incrível sabermos quem são os funcionários que irão pedir demissão no próximo mês. Nosso objetivo é saber disso 1 mês antes, para que possamos nos organizar e, além de chamar o funcionário para um papo aberto, podermos oferecer oportunidades que o motive a continuar conosco.





Neste primeiro cenário, que tipo de modelo o cientista de dados deveria buscar? Um modelo que decide, ordena ou estima?

Veja que na nossa missão como diretor de RH não existem restrições. Vamos atuar em todos os profissionais que forem pedir demissão. Sendo assim, o modelo que o cientista de dados deveria buscar é o modelo que decide corretamente. Ou seja, modelo de decisão.

Agora vamos imaginar que como diretores de RH estamos focados, ainda, na retenção de nossos funcionários. Queremos saber quem são os funcionários que irão pedir demissão no próximo mês, no entanto temos uma verba restrita para promoções (aumentos salariais) e como uma de nossas analistas está de férias, não teríamos recursos disponíveis para muitas intervenções (chamar o funcionário para um papo maroto).

Como estamos limitados em 3 intervenções mensais, precisamos da ajuda de um cientista de dados para que ele nos oriente quem são esses funcionários que agiremos.

Repare agora, que por mais que o objetivo ainda seja o mesmo, o cenário dois está muito diferente do cenário um. Neste momento, nós - diretores de RH - temos algumas limitações... Então, não adianta o cientista de dados dizer quem irá pedir demissão porque se essa quantidade for maior que a nossa capacidade, não saberemos em quem agir. Sendo assim, nesse cenário a utilidade do modelo muda para ordenação!

(um comentário entre parênteses: aqui não estamos questionando a importância do funcionário para o negócio, ok? Caso isso também deva ser levado em consideração, outras análises precisarão serem feitas a fim de agregar valor ao modelo de propensão de pedido de demissão.)

Veja que interessante, o contexto nos fez mudar a forma de utilização de um modelo! Eu acho isso muito animal...

Para provarmos a necessidade de um modelo com objetivo de estimativa, vamos criar um terceiro cenário dentro do nosso mundo de RH. Imagine que nossa empresa está passando por “apuros” financeiros e existe a ordem de cortes. Devemos reduzir nosso gasto com folha de pagamento e agora estamos muito preocupados, porque essas reduções causam uma grande tensão no ambiente de trabalho, principalmente quando o motivo é redução de custos!

Como sabemos que analiticamente é possível estimar a probabilidade dos meus funcionários pedirem demissão no mês seguinte e como estamos focados na redução de custo e qualidade do ambiente de trabalho, nós como ótimos diretores de RH, vamos solicitar este modelo para sabermos qual a probabilidade estimada de cada funcionário pedir demissão por livre e espontânea vontade. Esta probabilidade será multiplicada pelo salário do indivíduo, para sabermos quanto teríamos de economia no mês seguinte. Caso esta economia já bata nossa meta de redução de gastos com folha, deixaremos esses funcionários saírem, sem retê-los. Sendo assim, manteremos a qualidade do ambiente de trabalho (por não termos demitido ninguém) e ainda economizaremos com as rescisões.

Note que nesse novo cenário o propósito do modelo mudou completamente. O objetivo real do diretor de RH é estimar a redução de custo que ele terá. Sendo assim não tem interesse em saber quem e nem qual é o mais propenso.

Neste cenário o desafio do cientista de dados é criar estimativas super precisas e acuradas, para garantir que a conta posterior fará sentido e representará, realmente, a redução média, no próximo mês, de gasto em folha.

Ficaram claras as 3 utilidades de um modelo? E que não tem como não interagir com negócio para conseguir atingir o máximo proveito do trabalho de um cientista de dados?





04.

PAPEL DO DATA
SCIENTIST



PAPEL DO DATA SCIENTIST

Como argumentei anteriormente, não faz sentido um cientista de dados não interagir com a área que utilizará seus algoritmos. Por quê?

Porque ele perderá a capacidade de escolher métricas e, principalmente, focar no que importa para o negócio. Se um cientista de dados não se mantém alinhado com a necessidade, infelizmente, seus algoritmos não oferecerão os melhores resultados, causando em alguns casos, a frustração da área de negócio.

Na minha visão, o papel do cientista de dados não se limita à criação de um algoritmo ou outro. Ao meu ver o papel do cientista de dados é oferecer soluções reais e cabíveis no contexto no qual se está inserido. Sendo assim, o cientista de dados deve ser capaz de entender o problema pelas palavras de um leigo analiticamente. Além de estruturar a solução do problema, executar, entregar e explicar claramente o que o algoritmo representa e qual o real valor que ele agrega.





Se eu pudesse resumir o papel do cientista de dados, eu diria que é o cara que entende a dor de um leigo analiticamente; é capaz de traduzir o problema em dados, consegue manusear as informações disponíveis; é criativo o suficiente para criar novos dados a partir dos dados existentes e das conversas com o time de negócios; além de ter conhecimento de algoritmos que os suportem para a criação da solução.

Além disso tudo, o cientista de dados é um contador de histórias, ou seja, ele consegue fazer a apresentação de seus projetos, mostrando toda a sensualidade do estudo, destacando todos os benefícios que serão atingidos na utilização do mesmo.

É por isso que o cientista de dados, hoje em dia, é chamado de unicórnio. Algo raríssimo de se encontrar e que na maioria dos casos não se trata de apenas uma pessoa, mas um conjunto de pessoas trabalhando em grupo!



05.

CONCLUSÃO



CONCLUSÃO

Este e-book tem o objetivo de clarear os propósitos pelo qual um modelo é criado: decisão, ordenação e estimação.

Repare que para casos de previsões nominais (variáveis qualitativas - probabilidade de algo acontecer) os 3 tipos de modelos podem ser necessários, dependendo da história de negócio. Já quando estamos falando de previsões numéricas (variável resposta quantitativa) o único tipo de modelo que podemos desenvolver é o de estimação, pois sem dúvida, sempre vamos querer ser o mais precisos possível!

Espero que eu tenha sido clara o suficiente para lhe fazer sorrir em alguns momentos, trazendo verdades que lhe agreguem algum valor!

Se quiser bater algum papo, não deixe de me escrever... Como sempre digo para meus alunos: eu demoro, mas respondo!



ISENÇÃO DE RESPONSABILIDADE

Todas as informações contidas neste material são provenientes das minhas experiências pessoais ao longo dos anos. Embora eu tenha me esforçado ao máximo, para garantir a precisão e a mais alta qualidade dessas informações e que todas as técnicas e métodos aqui ensinados sejam altamente efetivos, eu não me responsabilizo por erros (ainda mais de português) ou omissões. Você deverá utilizar e ajustar as informações deste material de acordo com sua situação e necessidades.

DIREITOS AUTORAIS

Este material está protegido por leis de direitos autorais. Todos os direitos são reservados. Você não tem permissão para vender este material nem para copiar/reproduzir o conteúdo em sites, blogs, jornais ou quaisquer outros veículos de distribuição e mídia. Qualquer tipo de violação dos direitos autorais estará sujeita a ações legais.

Conforme a Lei 9.610/98, é proibida a reprodução total e parcial ou divulgação comercial sem a autorização prévia e expressa do autor (artigo 29).

Imagens ilustrativas por [freepik](#).

DIVIRTA-SE APRENDENDO DE VERDADE!



www.asn.rocks



Sobre Adriana Silva

Graduada em Estatística pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp – 2008), Mestre em Ciências: Estatística e Experimentação Agronômica pela Universidade de São Paulo (ESALQ/USP – 2012). Autora do livro “Data Mining Techniques to Acquire New Customers”, coordenadora e docente no curso “SAS Academy Data Science” do SAS, além de professora de Estatística e Machine Learning em trabalhos voluntários.

Vice Presidente eleita do CONRE-3 em 2019 e Analytics Director na Kantar Brasil. Já atuou como Head de Analytics no SAS, foi gerente analítica na unidade de negócio ABD (Abril Big Data) na Editora Abril e pré vendas na Oracle.

Tem como lema de vida o compartilhamento do pouco que sabe. Acredita que a ciência irá mudar o mundo e que todos tem capacidade de se divertir analisando dados, uma vez que seja bem feito!



[/in/adrianamms](https://www.linkedin.com/in/adrianamms)



dri@asn.rocks



www.asn.rocks



Como utilizar um modelo

Quando se cogita a criação de um modelo é importantíssimo saber para quê e como este modelo será utilizado. Este e-book detalha as 3 razões comuns da utilização de um modelo de machine learning.

Data: janeiro/2019

Divirta-se aprendendo de verdade!

www.asn.rocks