

Comunidade Data Science

(<https://www.comunidadedatascience.com/>)

Aumente o Faturamento da Empresa com Previsão de Churn

Introdução

Eu tenho uma boa e uma má notícia pra você. A má notícia é que a sua faculdade não é relevante para a carreira de Ciência de Dados. E a boa notícia é que a sua faculdade não é relevante para a carreira de Ciência de Dados.

Todas as grandes empresas, que entendem exatamente como um Cientista de Dados gerar valor para o negócio irá considerar sua faculdade.

O que as empresas buscam nos candidatos para vagas de Ciência de Dados, é sua capacidade de resolver problemas de negócio usando teorias, métodos e ferramentas da Ciência de Dados. E a forma mais simples e direta de demonstrar isso é construir um portfólio de projetos matador.

Leia até o final esse post, onde eu vou propor um problema de negócio para você criar uma solução usando os conhecimentos que você tem até o momento sobre Ciência de Dados.

Os 2 tipos de projetos em Data Science.

Existem 2 grandes tipos de projetos em Ciência de Dados: **Os projetos de Estudos e os projetos de Portfólio.**

Os projetos de Estudos são menos complexos e tem como objetivo fixar o aprendizado nos estudos, seja por meio da confirmação de algum conceito, a observação do comportamento de aprendizado de algum modelo de Machine Learning, a comprovação do resultado de uma técnica estatística, etc.

Os projetos de Portfólio são os mais complexos e tem como objetivo comprovar seu conhecimento em Ciência de dados. Esse tipo de projeto é muito próximo de um problema real, e geralmente não tem gabarito, ou seja, você precisa usar todo o conhecimento que você tem até aquele momento para criar uma solução para esse problema.

Infelizmente, muitos curso de Ciência de Dados, ensinam a profissão somente com projetos de Estudos e não validam o conhecimento dos alunos por meio de projetos desafiadores que são os projetos de Portfólio. Isso acontece por falta de conhecimento desses tipos de projetos por meios dos instrutores, que geralmente aprenderam Ciência de Dados somente através dos livros e nunca estiveram em um campo de batalha de verdade, dentro de uma empresa famosa, gerando valor para o negócio através da Ciência de Dados.

Para saber mais sobre os tipos de projetos, eu gravei um vídeo para o canal "**Seja um Data Scientist**", <https://youtu.be/LJrK4B7bNWA> (<https://youtu.be/LJrK4B7bNWA>), onde eu explico detalhadamente os 5 tipos de projetos obrigatórios para o seu portfólio.

E para te ajudar na criação do seus Projetos de Portfólio, eu vou sugerir um desafio real de Ciência de Dados com um contexto de negócio para você se orientar e também uma recomendação sobre como você deveria resolvê-lo.

A Empresa TopBank

Disclaimer: O Contexto a seguir, é completamente fictício, a empresa, o contexto, o CEO, as perguntas de negócio existem somente na minha imaginação.

A TopBank é uma grande empresa de serviços bancários. Ela atua principalmente nos países da Europa oferecendo produtos financeiros, desde contas bancárias até investimentos, passando por alguns tipos de seguros e produto de investimento.

O modelo de negócio da empresa é do tipo serviço, ou seja, ela comercializa serviços bancários para seus clientes através de agências físicas e um portal online.

O principal produto da empresa é uma conta bancária, na qual o cliente pode depositar seu salário, fazer saques, depósitos e transferência para outras contas. Essa conta bancária não tem custo para o cliente e tem uma vigência de 12 meses, ou seja, o cliente precisa renovar o contrato dessa conta para continuar utilizando pelos próximos 12 meses.

Segundo o time de Analytics da TopBank, cada cliente que possui essa conta bancária retorna um valor monetário de 15% do valor do seu salário estimado, se esse for menor que a média e 20% se esse salário for maior que a média, durante o período vigente de sua conta. Esse valor é calculado anualmente.

Por exemplo, se o salário mensal de um cliente é de R\$1.000,00 e a média de todos os salários do banco é de R\$800. A empresa, portanto, fatura R\$ 200 anualmente com esse cliente. Se esse cliente está no banco há 10 anos, a empresa já faturou R\$2.000,00 com suas transações e utilização da conta.

Nos últimos meses, o time de Analytics percebeu que a taxa de clientes cancelando suas contas e deixando o banco, atingiu números inéditos na empresa. Preocupados com o aumento dessa taxa, o time planejou um plano de ação para diminuir taxa de evasão de clientes.

Preocupados com a queda dessa métrica, o time de Analytics da TopBottom, contratou você como consultor de Data Science para criar um plano de ação, com o objetivo de reduzir a evasão de clientes, ou seja, impedir que o cliente cancele seu contrato e não o renove por mais 12 meses. Essa evasão, nas métricas de negócio, é conhecida como **Churn**.

De maneira geral, Churn é uma métrica que indica o **número de clientes que cancelaram o contrato ou pararam de comprar seu produto** em um determinado período de tempo. Por exemplo, clientes que cancelaram o contrato de serviço ou após o vencimento do mesmo, não renovaram, são clientes considerados em churn.

Outro exemplo seria os clientes que não fazem uma compra à mais de 60 dias. Esse clientes podem ser considerados clientes em churn até que uma compra seja realizada. O período de 60 dias é totalmente arbitrário e varia entre empresas.

O Desafio

Como um Consultor de Ciência de Dados, você precisa criar um plano de ação para diminuir o número de clientes em **churn** e mostrar o retorno financeiro da sua solução.

Ao final da sua consultoria, você precisa entregar ao CEO da TopBottom um modelo em produção, que receberá uma base de clientes via API e retornará essa mesma base "scorada", ou seja, um coluna à mais com a probabilidade de cada cliente entrar em churn.

Além disso, você precisará fornecer um relatório reportando a performance do seu modelo e o impacto financeiro da sua solução. Questões que o CEO e o time de Analytics gostariam de ver em seu relatório:

1. Qual a taxa atual de Churn da TopBank? Como ela varia mensalmente?
2. Qual a Performance do modelo em classificar os clientes como **churns**?
3. Qual o retorno esperado, em termos de faturamento, se a empresa utilizar seu modelo para evitar o **churn** dos clientes?

Uma possível ação para evitar que o cliente entre em **churn** é oferecer um cupom de desconto, ou alguma outro incentivo financeiro para ele renovar seu contrato por mais 12 meses.

1. Para quais clientes você daria o incentivo financeiro e qual seria esse valor, de modo a maximizar o ROI (Retorno sobre o investimento). Lembrando que a soma dos incentivos não pode ultrapassar os R\$10.000,00

Os Dados

O conjunto de dados que será utilizado para criar a solução para a TopBottom, está disponível na plataforma do Kaggle. Esse é o link: [ChurnDataset \(https://www.kaggle.com/mervetorkan/churndataset\)](https://www.kaggle.com/mervetorkan/churndataset).

Cada linha representa um cliente e cada coluna contém alguns atributos que descrevem esse cliente. O conjunto de dados inclui informações sobre:

- **RowNumber:** O número da coluna
- **CustomerID:** Identificador único do cliente
- **Surname:** Sobrenome do cliente.
- **CreditScore:** A pontuação de Crédito do cliente para o mercado de consumo.
- **Geography:** O país onde o cliente reside.
- **Gender:** O gênero do cliente.
- **Age:** A idade do cliente.
- **Tenure:** Número de anos que o cliente permaneceu ativo.
- **Balance:** Valor monetário que o cliente tem em sua conta bancária.
- **NumOfProducts:** O número de produtos comprado pelo cliente no banco.
- **HasCrCard:** Indica se o cliente possui ou não cartão de crédito.
- **IsActiveMember:** Indica se o cliente fez pelo menos uma movimentação na conta bancário dentro de 12 meses.
- **EstimateSalary:** Estimativa do salário mensal do cliente.
- **Exited:** Indica se o cliente está ou não em Churn.

Como Solucionar esse Desafio?

Não se assuste com o problema, respire fundo, mantenha a mente clara e limpa e então, comece a pensar de forma estruturada em alternativas para responder à essas perguntas.

Não tenha medo de criar suposições e considerações, faça um passo de cada vez, não se prenda muito na parte técnica e foque em responder as perguntas, todas suas ações devem te deixar um passo mais próximo da solução final. Sempre pense: "Se eu fizer isso, me ajuda a chegar mais próximo da resposta?" Se a resposta for Sim, faça, se não, tome outra ação.

Eu vou deixar aqui um roteiro para você se orientar, ele pode ser modificado da forma que você preferir ou simplesmente ignorado. Provavelmente, você já tem um roteiro de resolução melhor para abordar esse desafio.

A solução para esse desafio está disponível dentro da “Comunidade DS”, ao fazer parte da comunidade, você terá acesso à todas as resoluções dos problemas propostos aqui no Blog, Lives de Aprofundamento da teoria de Machine Learning, poderá participar de discussões técnicas e simulações de entrevistas, além de cursos específicos sobre ferramentas de DS e Clouds.

Para participar, você precisa se inscrever nesse link, [inscricao-lives-comunidade-DS – Seja Um Data Scientist \(https://sejaumdatascientist.com/inscricao-lives-comunidade-ds/\)](https://sejaumdatascientist.com/inscricao-lives-comunidade-ds/) e comparecer às Lives que acontece ao vivo no Youtube. Ao final da Live, eu abro vagas limitadas para quem deseja investir corretamente na carreira e participar da “Comunidade DS”.

Fica de olho no meu Instagram, @meigarom.datascience, onde eu divulgo a data e o horário da próxima Live ao vivo no Youtube.

Roteiro Sugerido para a Solução:

Esse é o roteiro de resolução do desafio que eu sugiro:

1. Explore os Dados usando ferramentas e testes Estatísticos com o objetivo de encontrar inconsistências dos dados e tratar possíveis dados faltantes.
2. Levante Hipóteses sobre as características de transações legítimas e fraudulentas, valide ou refute essas hipóteses através dos dados.
3. Prepare os Dados para que os Algoritmos Estatísticos ou de Machine Learning sejam capazes de aprender a tarefa.
4. Escreva vários algoritmos Classificadores.
5. Contabilize a Performance do Modelo de Machine Learning e transforme em Performance de Negócio.
6. Desenvolva uma API que retorne um “Score de Churn” para cada cliente, ou seja, um probabilidade do cliente entrar em churn.
7. Crie um README sobre como usar a sua solução.
8. Escreva um Relatório, respondendo as Perguntas do CEO da TopBottom.

O Ferramental da Solução

Use as ferramentas que você se sente mais confortável para desenvolver a solução. Você pode usar tanto Python quanto R e qualquer IDE de sua preferência Jupyter Notebook, Spyder, VS Code, entre outros.

Você pode usar o Google Colab também, caso você não tenha um computador razoável, ou caso queira testar essa incrível ferramenta do Google.

Aproveite esse projeto para melhorar sua velocidade na manipulação de dados com linguagens de programação. Alcance um nível, no qual você consiga escrever códigos rapidamente, sem ficar olhando no Stackoverflow a cada linha código.

Vá em Frente!

Não existe caminho fácil, de curto prazo em nenhuma profissão, muito menos em Data Science, mas existe o caminho certo. E o caminho certo é adquirir experiência através do desenvolvimento de projetos para mostrar sua capacidade.

Volto a repetir, os projetos do seu portfólio precisam demonstrar que você é tão capaz de resolver desafios de negócio quantos os Data Scientists que já atuam profissionalmente nas empresas.

Quando você solucionar esse desafio, escreva um artigo, explicando toda sua linha de raciocínio ou se inscreva na "Consultoria Data Science" através desse link: <https://sejaumdatascientist.com/inscricao-consultoria-ds/> (<https://sejaumdatascientist.com/inscricao-consultoria-ds/>) e eu pessoalmente revisarei sua solução e indicarei melhorias.

Se quiser publicar aqui no blog, me manda um msg no LinkedIn (@meigarom) ou no Instagram (@meigarom.datascience). Publicarei seu trabalho com o maior prazer do mundo.

Conclusão:

Nesse post, você recebeu um desafio de Ciência de Dados muito próximo dos desafios reais das empresas. Na vida real, os problemas chegam em forma de perguntas abertas, desestruturadas e sem nenhuma dica sobre como resolver. É papel do Data Scientist entender a causa raiz, planejar o desenvolvimento e criar a melhor solução para o problema de negócio verdadeiro.

Não se esqueça de acompanhar o canal "Seja um Data Scientist" e o Instagram @meigarom.datascience para mais conteúdos. Caso você tenha LinkedIn, não se esqueça de conectar comigo, é só procurar por Meigarom, está fácil de encontrar!! haha.

Bons estudos!!

Com a tag [ciência de dados](https://www.comunidade.datascience.com/tag/ciencia-de-dados/) (<https://www.comunidade.datascience.com/tag/ciencia-de-dados/>) [Cientista de Dados](https://www.comunidade.datascience.com/tag/cientista-de-dados/) (<https://www.comunidade.datascience.com/tag/cientista-de-dados/>) [Data Science](https://www.comunidade.datascience.com/tag/data-science/) (<https://www.comunidade.datascience.com/tag/data-science/>) [Data Scientist](https://www.comunidade.datascience.com/tag/data-scientist/) (<https://www.comunidade.datascience.com/tag/data-scientist/>) [desafio de classificacao](https://www.comunidade.datascience.com/tag/desafio-de-classificacao/) (<https://www.comunidade.datascience.com/tag/desafio-de-classificacao/>) [Machine Learning](https://www.comunidade.datascience.com/tag/machine-learning/) (<https://www.comunidade.datascience.com/tag/machine-learning/>) [métricas de churn](https://www.comunidade.datascience.com/tag/metricas-de-churn/) (<https://www.comunidade.datascience.com/tag/metricas-de-churn/>) [portfolio de projetos](https://www.comunidade.datascience.com/tag/portfolio-de-projetos/) (<https://www.comunidade.datascience.com/tag/portfolio-de-projetos/>) [previsao de Churn](https://www.comunidade.datascience.com/tag/previsao-de-churn/) (<https://www.comunidade.datascience.com/tag/previsao-de-churn/>).

Deixe um comentário

O seu endereço de e-mail não será publicado. Campos obrigatórios são marcados com *

Comentário *

Nome *

E-mail *

Site

☐ Salvar meus dados neste navegador para a próxima vez que eu comentar.

Publicar comentário

Todos os direitos reservados