


Comunidade Data Science

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

Como usar Data Science para fazer a empresa vender mais?

 Nesse artigo, eu vou mostrar um projeto muito utilizado por empresas para aumentar a receita através da oferta mais inteligente dos seus próprios produtos.

Ícones feitos por *Flat Icons* (<https://flat-icons.com/>) from *www.flaticon.com* (<https://www.flaticon.com/br/>).

Introdução

Se você chegou nesse artigo, você está procurando um projeto relevante para construir uma solução usando seus conhecimentos em Ciência de Dados e assim adicioná-lo ao seu portfólio de projetos.

Nesse artigo, eu vou mostrar um projeto muito utilizado por empresas para aumentar a receita através da oferta mais inteligente dos seus próprios produtos.

E para tornar esse projeto o mais próximo da realidade, eu vou criar um contexto de negócio e perguntas que o time de negócio faria pra você durante a reunião de início de projeto.

Ao responder essas perguntas de negócio, o seu projeto deixa de ser apenas uma prática de Machine Learning, para se tornar uma solução de um desafio real de negócio.

E é exatamente isso que faz seu projeto ser mais atrativo para os recrutadores.

O que um projeto de Data Science deve conter para ser atrativo?

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

Os projetos de Estudos tem o objetivo de fixar o aprendizado, confirmar um conceito aprendido ou avaliar seu conhecimento atual. Geralmente, são projetos mais simples e diretos, que não envolvem questões de negócio.

Os projetos de Portfólio, por outro lado, tem o objetivo de demonstrar seu conhecimento em Ciência de Dados, ou seja, ele mostra exatamente o que você faria para resolver os problemas de negócio, se estivesse dentro da empresa, como parte do time de Ciência de Dados.

Para saber mais sobre as etapas que são esperadas dos projetos de portfólio, assista esse vídeo <https://youtu.be/LJrK4B7bNWA> (<https://youtu.be/LJrK4B7bNWA>), onde eu explico detalhadamente os 5 tipos de projetos obrigatórios para o seu portfólio.

Para esse artigo, minha sugestão para aumentar a receita da empresa de maneira inteligente, usando os próprios produtos, é a seguinte:

A Empresa Insurance All

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

A Insurance All é uma empresa que fornece seguro de saúde para seus clientes e o time de produtos está analisando a possibilidade de oferecer aos assegurados, um novo produto: Um seguro de automóveis.

Assim como o seguro de saúde, os clientes desse novo plano de seguro de automóveis precisam pagar um valor anualmente à Insurance All para obter um valor assegurado pela empresa, destinado aos custos de um eventual acidente ou dano ao veículo.

A Insurance All fez uma pesquisa com cerca de 380 mil clientes sobre o interesse em aderir a um novo produto de seguro de automóveis, no ano passado. Todos os clientes demonstraram interesse ou não em adquirir o seguro de automóvel e essas respostas ficaram salvas em um banco de dados junto com outros atributos dos clientes.

O time de produtos selecionou 127 mil novos clientes que não responderam a pesquisa para participar de uma campanha, no qual receberão a oferta do novo produto de seguro de automóveis. A oferta será feita pelo time de vendas através de ligações telefônicas.

Contudo, o time de vendas tem uma capacidade de realizar 20 mil ligações dentro do período da campanha.

O Desafio

Nesse contexto, você foi contratado como um consultor de Ciência de Dados para construir um modelo que prediz se o cliente estaria ou não interessado no seguro de automóvel.

Com a sua solução, o time de vendas espera conseguir priorizar as pessoas com maior interesse no novo produto e assim, otimizar a campanha realizando apenas contatos aos clientes mais propensos a realizar a compra.

Como resultado da sua consultoria, você precisará entregar um relatório contendo algumas análises e respostas às seguintes perguntas:

1. Principais Insights sobre os atributos mais relevantes de clientes interessados em adquirir um seguro de automóvel.
2. Qual a porcentagem de clientes interessados em adquirir um seguro de automóvel, o time de vendas conseguirá contatar fazendo 20.000 ligações?
3. E se a capacidade do time de vendas aumentar para 40.000 ligações, qual a porcentagem de clientes interessados em adquirir um seguro de automóvel o time de vendas conseguirá contatar?
4. Quantas ligações o time de vendas precisa fazer para contatar 80% dos clientes interessados em adquirir um seguro de automóvel?

Os Dados

O conjunto de dados está disponível na plataforma do Kaggle, através desse link:

<https://www.kaggle.com/anmolkumar/health-insurance-cross-sell-prediction>
(<https://www.kaggle.com/anmolkumar/health-insurance-cross-sell-prediction>).

Cada linha representa um cliente e cada coluna contém alguns atributos que descrevem esse cliente, além da sua resposta à pesquisa, na qual ela mencionou interesse ou não ao novo produto de seguros.

O conjunto de dados inclui as seguintes informações:

- **Id:** identificador único do cliente.

- **Gender:** gênero do cliente.
- **Age:** idade do cliente.
- **Driving License:** 0, o cliente não tem permissão para dirigir e 1, o cliente tem para dirigir (CNH – Carteira Nacional de Habilitação)
- **Region Code:** código da região do cliente.
- **Previously Insured:** 0, o cliente não tem seguro de automóvel e 1, o cliente já tem seguro de automóvel.
- **Vehicle Age:** idade do veículo.
- **Vehicle Damage:** 0, cliente nunca teve seu veículo danificado no passado e 1, cliente já teve seu veículo danificado no passado.
- **Annual Premium:** quantidade que o cliente pagou à empresa pelo seguro de saúde anual.
- **Policy sales channel:** código anônimo para o canal de contato com o cliente.
- **Vintage:** número de dias que o cliente se associou à empresa através da compra do seguro de saúde.
- **Response:** 0, o cliente não tem interesse e 1, o cliente tem interesse.

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 a 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 será feita pelos membros da Comunidade DS sob a minha orientação e mentoria. Como conteúdo de apoio, todas as semanas acontecerá Lives de 1h30 de conteúdo aprofundado, onde ensinarei estratégias, métodos, modelos e interpretação de negócio necessários para criar uma ótima solução para esse problema.

Além disso, teremos Lives semanais de 1h para discutir o progresso dos membros, apontar melhorias, esclarecer dúvidas e discutir formas de resolução.

Ao fazer parte da Comunidade DS, você terá acesso a todas as resoluções dos projetos anteriores propostos aqui no Blog, junto com os códigos, notebooks, Lives de Aprofundamento e discussões gravadas.

Além disso, você terá acesso às formações base de um Cientista de Dados como Python, SQL, Storytelling, Machine Learning, Estatística, Engenharia de Software e Clouds.

A Comunidade abre vagas para novos membros de tempos em tempos. Fica de olho no meu Instagram, @meigarom.datascience, eu sempre trago informações relacionadas a Comunidade DS, além de responder perguntas e fornecer dicas de estudos.

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 Propensão" de compra para cada cliente, ou seja, uma probabilidade do cliente estar interessado em comprar um seguro de automóvel.
7. Crie um README sobre como interpretar e usar a sua solução.
8. Escreva um Relatório, respondendo às perguntas de negócio.

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!!

Meigarom

