**Os princípios básicos da engenharia de ML**

Agora que você tem uma ideia geral do que é a engenharia de ML, podemos nos concentrar um pouco nos elementos-chave que compõem essas categorias incrivelmente amplas na Figura 1.2. Cada um desses tópicos é o foco de discussões detalhadas sobre o comprimento de um capítulo posteriormente neste livro, mas por enquanto vamos olhar para eles em um sentido holístico por meio de cenários potencialmente dolorosamente familiares para elucidar por que eles são tão importantes. **Planejar**

*Nada é mais desmoralizante do que construir uma solução de ML que resolva o problema errado.*

De longe, a maior causa de falhas de projeto, não conseguir planejar completamente um projeto, é uma das maneiras mais desmoralizantes para um projeto ser cancelado. Imagine por um momento que você é o primeiro DS contratado de uma empresa. Na sua primeira semana, um executivo de marketing se aproxima, explicando (em seus termos) um problema sério de negócios que eles estão tendo. Eles precisam descobrir um meio eficiente de se comunicar com os clientes por e-mail para informar sobre as próximas vendas que eles podem estar interessados. Se estas são as únicas informações fornecidas e as consultas repetidas aos membros da equipe de marketing simplesmente declaram o mesmo objetivo final de aumentar a taxa de cliques e abertura, o número de vias a seguir parece ilimitada. Deixado por conta própria, você :

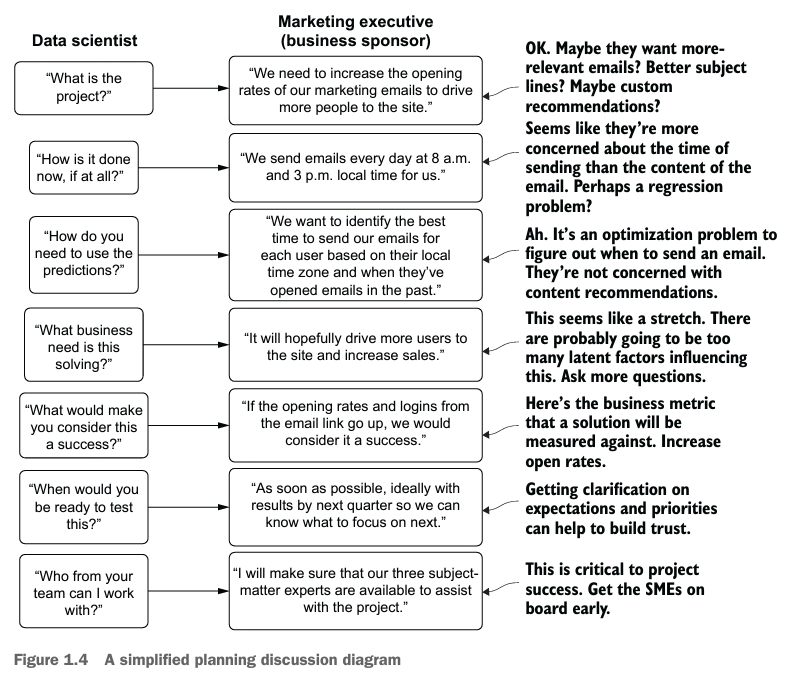
Focar na recomendação de conteúdo e cria e-mails personalizados para cada usuário?

Fornecer previsões com um sistema apoiado por NLP que criará linhas de assunto relevantes para cada usuário?

Tentar prever uma lista de produtos mais relevantes para a base de clientes para colocar à venda todos os dias?

Com tantas opções de complexidade e abordagens variadas, e pouca orientação, é altamente improvável criar uma solução alinhada com as expectativas do executivo.

Em vez disso, se uma discussão de planejamento adequada se aprofundar na quantidade correta de detalhes, evitando a complexidade do lado do ML das coisas, a verdadeira expectativa pode ser revelada. Você saberia então que a única expectativa é uma previsão de quando cada usuário provavelmente estaria aberto a ler e-mails. O executivo simplesmente quer saber quando é mais provável que alguém não esteja trabalhando, viajando ou dormindo para que a empresa possa enviar lotes de e-mails ao longo do dia para diferentes grupos de clientes. A triste realidade é que muitos projetos de ML começam dessa maneira. Freqüentemente, pouca comunicação ocorre em relação ao início do projeto, e a expectativa geral é que a equipe do DS apenas descubra. No entanto, sem a orientação adequada sobre o que precisa ser construído, como ele precisa funcionar e qual é o objetivo final das previsões, o projeto quase certamente está fadado ao fracasso. Afinal, o que teria acontecido se todo o sistema de recomendação de conteúdo fosse construído para esse caso de uso, com meses de desenvolvimento e esforço desperdiçados, quando uma simples consulta analítica baseada na geolocalização do endereço IP era o que realmente era necessário? O projeto não seria apenas cancelado, mas muitas questões provavelmente viriam de cima para saber por que esse sistema foi construído e por que seu desenvolvimento custou tanto. Vejamos a discussão de planejamento simplificado ilustrada na Figura 1.4. Mesmo na fase inicial de discussão, podemos ver como apenas algumas perguntas cuidadosas e respostas claras podem fornecer a única coisa que todo cientista de dados deve procurar nessa situação (especialmente como o primeiro DS em uma empresa que trabalha no primeiro problema): **uma vitória rápida**. Como você pode ver no monólogo interno do DS mostrado à direita, o problema em questão não está na lista de suposições originais feitas. Não se fala em conteúdo de e-mail, relevância para a linha de assunto ou itens do e-mail. É uma consulta analítica simples para descobrir em qual fuso horário os clientes estão e analisar a abertura histórica no horário local de cada cliente. Ao demorar alguns minutos para planejar e entender completamente o caso de uso, semanas (se não meses) de desperdício de esforço, tempo e dinheiro foram economizados. Ao se concentrar no que será construído e por que precisa ser construído, tanto a equipe do DS quanto a empresa podem orientar a discussão com mais fruição. Evitar uma conversa focada em como será construído mantém os membros do DS do grupo focados no problema. Ignorar quando será construído ajuda a empresa a manter seu foco alinhado nas necessidades do projeto. Evitar discutir os detalhes da implementação nesta fase do projeto não é apenas fundamental para a equipe se concentrar no problema. Manter os detalhes esotéricos de algoritmos e design de solução fora das discussões com a equipe maior mantém os membros da unidade de negócios engajados. Afinal, eles realmente não se importam com quantos ovos entram na mistura, de que cor são os ovos ou até mesmo de quais espécies põem os ovos; Eles só querem comer o bolo quando terminar. Abordaremos os processos de planejamento, conversando sobre expectativas de projetos com clientes empresariais internos e comunicações gerais sobre o trabalho de ML com um público não técnico em profundidade e com muito mais profundidade durante o restante da Parte 1.



Qual é o projeto?”

Precisamos aumentar as taxas de abertura de nossos e-mails de marketing para levar mais pessoas ao site.”

Certo. Talvez eles queiram e-mails mais relevantes? Melhores linhas de assunto? Talvez recomendações personalizadas?

“Como é feito agora, se é que está acontecendo?”

“Enviamos e-mails todos os dias às 8h e 15h. hora local para nós.”

Parece que eles estão mais preocupados com o tempo de envio do que com o conteúdo do e-mail. Talvez um problema de regressão?

“Como você precisa usar as previsões?”

“Queremos identificar o melhor momento para enviar nossos e-mails para cada usuário com base em seu fuso horário local e quando eles abriram e-mails no passado.”

É um problema de otimização descobrir quando enviar um e-mail. Eles não estão preocupados com recomendações de conteúdo.

“Que necessidade de negócios isso está resolvendo?”

“Espero que isso leve mais usuários ao site e aumente as vendas.”

Isso parece um exagero. Provavelmente haverá muitos fatores latentes que influenciam isso. Faça mais perguntas.

“O que faria você considerar isso um sucesso?”

“Se as taxas de abertura e os logins do link do e-mail aumentarem, consideraremos um sucesso.”

Aqui está a métrica de negócios com a qual uma solução será medida. Aumente as taxas de abertura.

“Quando você estaria pronto para testar isso?”

“O mais rápido possível, idealmente com resultados no próximo trimestre para que possamos saber no que focar a seguir.”

Obter esclarecimentos sobre expectativas e prioridades pode ajudar a construir confiança.

“Com quem da sua equipe posso trabalhar?”

“Vou garantir que nossos três especialistas em assuntos estejam disponíveis para ajudar no projeto.”

Isso é fundamental para o sucesso do projeto. Leve as PMEs a bordo cedo.