3.1.1 Planejamento básico para um projeto

O planejamento de qualquer projeto de ML normalmente começa em alto nível. Uma unidade de negócios, um executivo ou até mesmo um membro da equipe DS tem a ideia de usar a experiência da equipe Data Science para resolver um problema desafiador. Embora normalmente seja pouco mais que um conceito nesta fase inicial, este é um momento crítico no ciclo de vida de um projeto. No cenário que estamos discutindo, a ideia de alto nível é a personalização. Para um DS experiente, isso pode significar uma série de coisas. Para uma PME da unidade de negócios, isso poderia significar muitos dos mesmos conceitos que a equipe de DS poderia imaginar, mas pode não ser. Desde o início de uma ideia até antes mesmo do início da pesquisa básica, a primeira coisa que todos os envolvidos neste projeto deveriam fazer é realizar uma reunião. O tema desta reunião deverá centrar-se num elemento fundamental: Por que estamos construindo isto? Pode parecer uma pergunta hostil ou de confronto. Pode surpreender algumas pessoas ao ouvi-lo. No entanto, é uma das questões mais eficazes e importantes, pois abre uma discussão sobre as verdadeiras motivações pelas quais as pessoas desejam que o projeto seja construído. É para aumentar as vendas? É para deixar nossos clientes externos mais felizes? Ou é para manter as pessoas navegando no site por mais tempo? Cada uma dessas respostas diferenciadas pode ajudar a informar o objetivo desta reunião: definir as expectativas do resultado de qualquer trabalho de ML. A resposta também satisfaz os critérios da métrica de medição do desempenho do modelo, bem como a pontuação de atribuição do desempenho na produção (a mesma pontuação que será usada para medir os testes A/B muito mais tarde). Em nosso cenário de exemplo, a equipe não consegue fazer essa importante pergunta do porquê. A Figura 3.6 mostra a divergência nas expectativas do lado empresarial e do lado ML porque nenhum dos grupos está falando sobre o aspecto essencial do projeto e, em vez disso, está ocupado em silos mentais que eles próprios criaram. A equipe de ML está focada inteiramente em como resolver o problema, enquanto a equipe de negócios tem expectativas sobre o que seria entregue, presumindo erroneamente que a equipe de ML “simplesmente entenderá”. A Figura 3.6 resume o processo de planejamento do MVP. Com requisitos extremamente vagos, uma completa falta de comunicação completa sobre as expectativas quanto à funcionalidade mínima do protótipo e uma falha em controlar a complexidade da experimentação, a demonstração é considerada um fracasso absoluto. A prevenção de resultados como este só pode ser alcançada nestas primeiras reuniões, quando as ideias do projecto estão a ser discutidas. Ampliar a sobreposição entre essas regiões de lacunas de expectativas é responsabilidade do líder da equipe de DS e do gerente do projeto. Na conclusão das reuniões de planejamento, um estado ideal é o alinhamento das expectativas de todos (sem que ninguém se concentre em detalhes de implementação ou em funcionalidades específicas fora do escopo a serem potencialmente adicionadas no futuro). Continuando com este cenário, vamos dar uma olhada na discussão de feedback da demonstração do MVP para ver os tipos de questões que poderiam ter sido discutidas durante aquela reunião inicial de planejamento e definição do escopo. A Figura 3.7 mostra as questões e as causas subjacentes dos actuais mal-entendidos.

Embora este exemplo seja intencionalmente hiperbólico, encontrei elementos dessa confusão presentes em muitos projetos de ML (aqueles fora de empresas focadas principalmente em ML), e isso é de se esperar. Os problemas que o ML frequentemente pretende resolver são complexos, cheios de detalhes específicos e únicos de cada negócio (e unidade de negócios dentro de uma empresa) e repletos de desinformação em torno das mínimas nuances desses detalhes. É importante perceber que essas lutas serão uma parte inevitável de qualquer projeto. A melhor maneira de minimizar seu impacto é realizar uma série de discussões completas que visam capturar o máximo possível de detalhes sobre o problema, os dados e as expectativas do resultado.

