**Planejamento: Você quer que eu preveja o quê?!**

Antes de entrarmos em como são realizadas as fases de planejamento bem-sucedidas de projetos de ML, vamos fazer uma simulação da gênese de um projeto típico em uma empresa que não possui um processo estabelecido ou comprovado para iniciar o trabalho de ML. Vamos imaginar que trabalhamos em uma empresa de e-commerce que está começando a sentir o gostinho de querer modernizar seu site. Depois de ver concorrentes anunciando enormes ganhos de vendas ao adicionar serviços de personalização a seus sites durante anos, a equipe de nível C está exigindo que a empresa siga todas as recomendações. Ninguém no alto escalão tem certeza dos detalhes técnicos sobre como esses serviços são construídos, mas todos sabem que o primeiro grupo com quem conversar são os nerds de ML. A empresa (neste caso, a liderança do departamento de vendas, marketing e equipes de produto) convoca uma reunião, convidando toda a equipe de ML, com pouca cor adicional ao convite além do título “Início do projeto de recomendações personalizadas”. A gerência e os vários departamentos com os quais você trabalhou ficaram satisfeitos com os projetos de ML de pequena escala que sua equipe construiu (detecção de fraude, estimativa de avaliação de clientes, previsão de vendas e modelos de risco de probabilidade de rotatividade). Cada um dos projetos anteriores, embora complexos em vários aspectos do ponto de vista de ML, eram em grande parte isolados – conduzidos pela equipe de ML, que apresentou uma solução que poderia ser consumida pelas diversas unidades de negócios. Nenhum desses projetos exigiu estimativas subjetivas de qualidade ou regras comerciais excessivas para influenciar os resultados. A pureza matemática destas soluções simplesmente não estava aberta a argumentos ou interpretações; ou eles estavam certos ou estavam errados. Vítima do seu próprio sucesso, a equipe é abordada pelo negócio com um novo conceito: modernizar o site e os aplicativos mobile. Os executivos ouviram falar dos enormes ganhos de vendas e da fidelização de clientes que acompanham as recomendações personalizadas e querem que sua equipe construa um sistema para incorporação ao site e aos aplicativos. Eles querem que cada usuário veja uma lista exclusiva de produtos que os cumprimentam quando fazem login. Eles querem que esses produtos sejam relevantes e interessantes para o usuário e, no final das contas, querem aumentar as chances de que o usuário comprará esses itens. Após uma breve reunião durante a qual são mostrados exemplos de outros sites, eles perguntam quanto tempo levará até que o sistema esteja pronto. Você estima cerca de dois meses, com base nos poucos artigos que leu no passado sobre esses sistemas, e começa a trabalhar. A equipe cria um plano de desenvolvimento provisório durante a próxima reunião scrum e todos partem para tentar resolver o problema. Você e o restante da equipe de ML presumem que a gestão está buscando o comportamento mostrado em tantos outros sites, nos quais os produtos são recomendados em uma tela principal. Afinal, isso é personalização no seu sentido mais puro: uma coleção única de produtos que um algoritmo previu que terá relevância para um usuário individual. Essa abordagem parece bastante simples, todos vocês concordam, e a equipe começa a planejar rapidamente como construir um conjunto de dados que mostre uma lista classificada de chaves de produto para cada usuário do site e do aplicativo móvel, com base apenas no histórico de navegação e compra de cada membro.

Nos próximos sprints, todos vocês trabalharão cuidadosamente de forma isolada. Você testa dezenas de implementações que viu em postagens de blog, consome centenas de artigos teóricos sobre diferentes algoritmos e abordagens para resolver um problema de recomendação implícita e, finalmente, constrói uma solução MVP usando mínimos quadrados alternados (ALS) que atinge uma raiz do erro quadrático médio (RMSE) de 0,2334, juntamente com uma implementação aproximada de pontuação ordenada para relevância com base no comportamento anterior. Cheio de confiança de que você tem algo incrível para mostrar ao patrocinador da equipe de negócios, você vai para a reunião munido do caderno de testes, gráficos que mostram as métricas gerais e exemplos de dados de inferência que você acredita que realmente impressionarão a equipe. Você começa mostrando a pontuação geral em escala para afinidade, exibindo os dados como um gráfico RMSE, conforme mostrado na figura 3.3.

A resposta ao mostrar o gráfico é, na melhor das hipóteses, morna. Surgem muitas questões focadas no que os dados significam, no que significa a linha que cruza os pontos e como os dados foram gerados. Em vez de uma discussão focada sobre a solução e a próxima fase na qual você gostaria de trabalhar (aumentando a precisão), a reunião começa a se transformar em uma mistura de confusão e tédio. Em um esforço para explicar melhor os dados, você mostra uma tabela rápida de eficácia de classificação usando métricas de ganho cumulativo não descontado (NDCG) para ilustrar o poder preditivo de um único usuário escolhido aleatoriamente, conforme mostrado na figura 3.4.

O primeiro gráfico criou uma leve sensação de perplexidade, mas a tabela traz confusão total e completa. Ninguém entende o que está sendo mostrado ou consegue ver a relevância para o projeto. A única coisa que passa pela cabeça de todos é: “É realmente isso que semanas de esforço podem trazer? O que a equipe de ciência de dados tem feito todo esse tempo?” Durante a explicação das duas visualizações pela equipe de DS, um dos analistas de marketing começa a pesquisar a lista de recomendações de produtos para as contas de um dos membros da equipe no conjunto de dados de amostra fornecido para a reunião. A Figura 3.5 ilustra os resultados juntamente com os pensamentos do analista de marketing ao trazer os dados do catálogo de produtos para cada recomendação da lista. A maior lição que a equipe do DS tira desta reunião não é, de fato, a necessidade de validar os resultados do seu modelo de uma forma que simule a forma como um usuário final das previsões reagiria. Embora seja uma consideração importante, e discutida na barra lateral a seguir, ela é superada de forma bastante significativa pela percepção de que a razão pela qual o modelo foi recebido tão mal foi a de que a equipe não planejou adequadamente as nuances deste projeto.

A equipe do DS simplesmente não entendeu o problema de negócios da perspectiva dos outros membros da equipe presentes, que sabiam onde todos os proverbiais “corpos foram enterrados” nos dados e que acumulavam décadas de conhecimento sobre a natureza dos dados e do produto. O ônus dessa falha não recai apenas sobre o gerente de projeto, o líder da equipe de DS ou qualquer membro da equipe. Pelo contrário, trata-se de uma falha coletiva de cada membro da equipe mais ampla ao não definir minuciosamente o âmbito e os detalhes do projeBto. Como eles poderiam ter feito as coisas de forma diferente? O analista que consultou suas próprias previsões descobriu muitos problemas que eram óbvios para eles. Eles viram os dados de itens duplicados devido à desativação de IDs de produtos mais antigos e também souberam instantaneamente que a divisão de calçados usava um ID de produto separado para cada cor de estilo de calçado, dois problemas principais que causaram uma demonstração ruim. Todos os problemas encontrados, que causam alto risco de cancelamento do projeto, foram devido ao planejamento inadequado do projeto.

**Espere um minuto. O planejamento de um projeto não está em desacordo com o Agile?**

Bem, sim e não. Para citar Scott Ambler (um dos escritores mais prolíficos sobre processos fundamentais para Agile): “Um plano de projeto é importante, mas não deve ser muito rígido para acomodar mudanças na tecnologia ou no ambiente, nas prioridades das partes interessadas e na compreensão das pessoas sobre o problema e sua solução” (http://www.ambysoft.com/essays/agileManifesto.html). Tenho visto a interpretação errônea desse sentimento surgir com bastante frequência em minha carreira. Ambler e os criadores originais do Manifesto Ágil apontavam que um projeto não deveria ser ditado por um roteiro pré-planejado e imutável de elementos que precisam ser construídos. A intenção não é, e nunca foi, não planejar nada. Trata-se simplesmente de ser flexível nos planos que se elaboram, para permitir que sejam alterados quando surgir a necessidade. Se surgir uma maneira mais simples de implementar algo, uma maneira melhor que reduza a complexidade e ao mesmo tempo alcance o mesmo resultado final, então o plano do projeto deverá mudar. No mundo do ML, esta é uma ocorrência frequente. Talvez, no início do projeto (antes de uma fase de pesquisa completa ser concluída), a equipe multifuncional determine que a única solução possível é uma abordagem de modelagem altamente complexa e complicada. Depois de realizar experimentos, entretanto, a equipe descobre que uma equação linear simples poderia ser desenvolvida para resolver o problema com precisão aceitável por uma fração do tempo e custo de desenvolvimento. Embora o plano inicial fosse usar, digamos, aprendizagem profunda para resolver o problema, a equipe pode, deve e deve mudar de direção para uma abordagem muito mais simples. O plano mudou, certamente, mas sem um plano em primeiro lugar, a fase de pesquisa e experimentação seria como um navio perdido na noite – sem guia, sem direção e movendo-se caoticamente no escuro. O planejamento é bom em ML. É fundamental não definir esses planos em pedra.

**Não confie cegamente em suas métricas**

Ao realizar ML particularmente em grande escala, confiar fortemente em métricas de erro e pontuações de validação para modelos é incrivelmente tentador. Eles não são apenas o único meio verdadeiramente realista de medir a qualidade objetiva de previsões em grandes conjuntos de dados (com os quais muitos de nós lidamos com frequência hoje em dia), mas também são frequentemente o único meio quantitativo real e válido de julgar a qualidade preditiva de um determinado implementação. No entanto, é importante não confiar apenas nessas métricas de pontuação do modelo. Use-os (isto é, os apropriados para o trabalho em questão), mas complemente-os com meios adicionais de obter medições subjetivas da eficácia da previsão. Conforme mostrado na figura 3.5, uma simples visualização das previsões para um usuário individual revelou muito mais avaliações de qualidade objetivas e subjetivas do que qualquer algoritmo preditivo de pontuação de ordenação ou estimativa de perda jamais poderia fazer. Tenha em mente que esta avaliação adicional da amostra de simulação de uso final não deve ser feita pelos membros da equipe de DS, a menos que eles estejam julgando a qualidade da previsão dos dados pelos quais eles próprios são considerados PMEs. Para o caso de uso que estamos discutindo, caberia à equipe de DS fazer parceria com alguns analistas de marketing para fazer uma validação informal de garantia de qualidade (QA) antes de mostrar os resultados para a equipe maior.