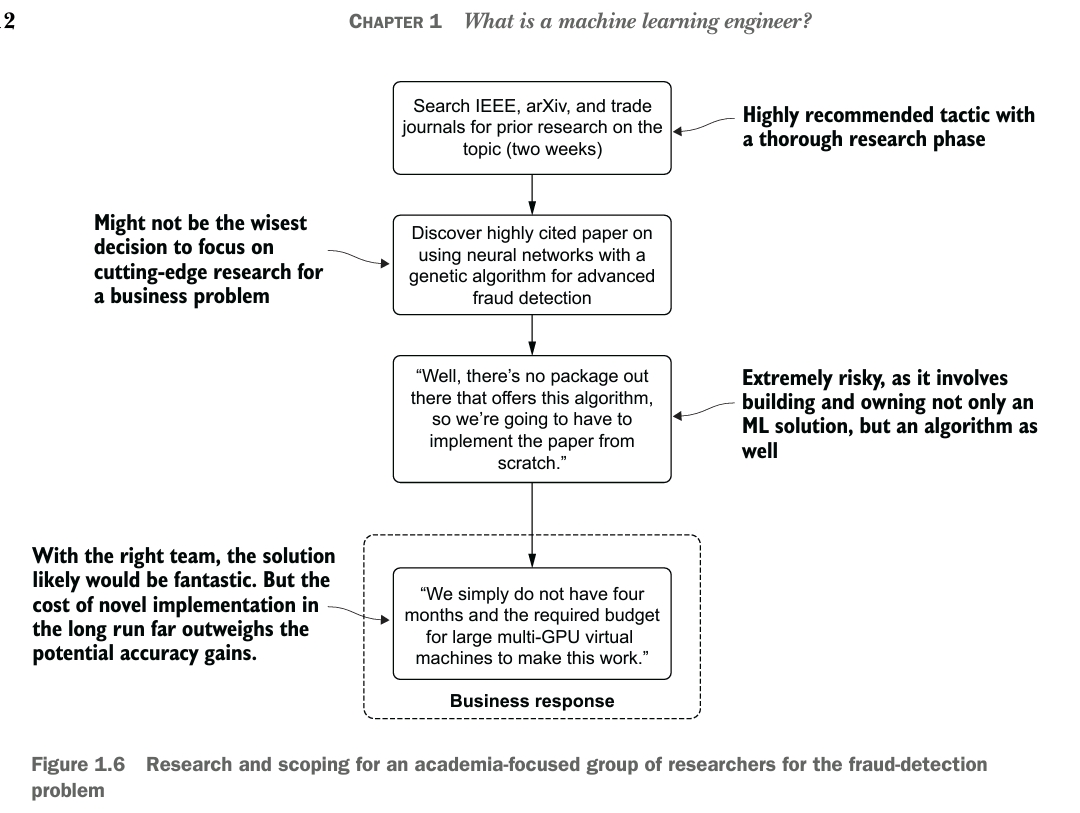
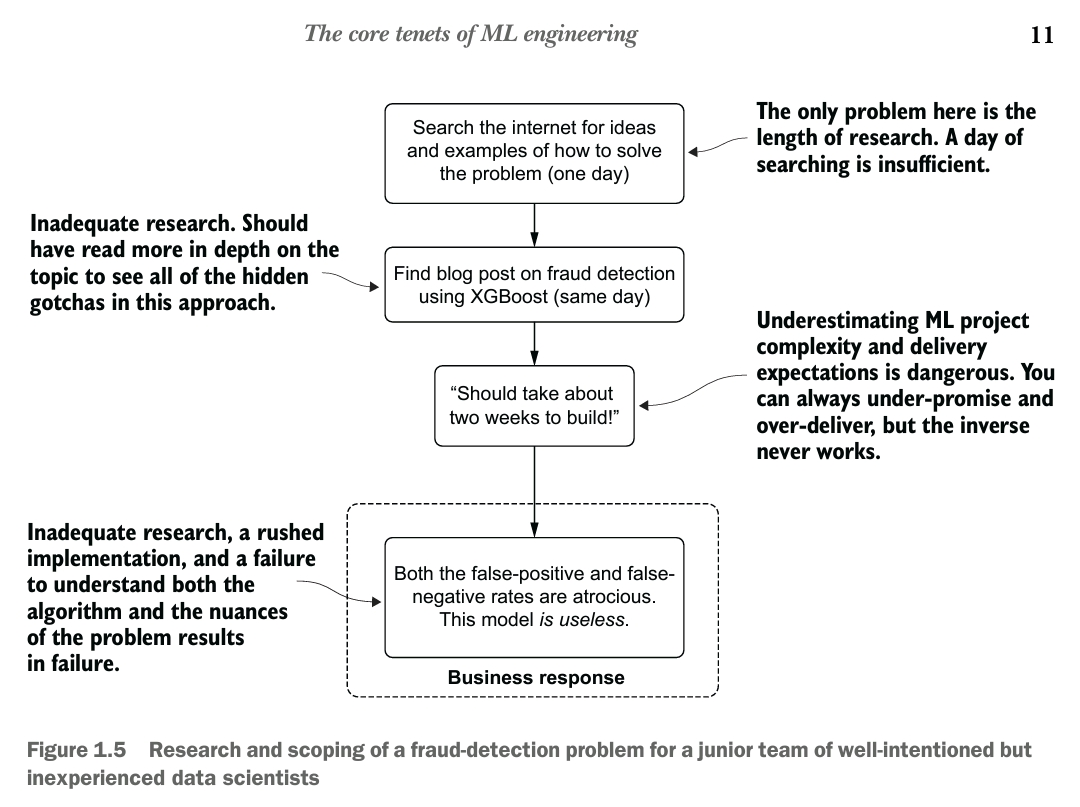
1.2.2 **Escopo e pesquisa**

Se você mudar sua abordagem no meio do desenvolvimento, enfrentará uma conversa difícil com a empresa para explicar que os atrasos do projeto se devem ao fato de você não ter feito sua lição de casa. Afinal, existem apenas duas dúvidas que seus clientes internos (a unidade de negócios) têm sobre o projeto: Isso vai resolver meu problema? Quanto tempo isso vai levar? Vamos dar uma olhada em outro cenário potencialmente familiar para discutir maneiras opostas pelas quais esse estágio de desenvolvimento de projetos de ML pode dar errado. Digamos que temos duas equipes de DS em uma empresa, cada uma enfrentando a outra para desenvolver uma solução para uma crescente incidência de fraudes conduzidas no sistema de faturamento da empresa. O processo de pesquisa e definição do escopo da Equipe A é ilustrado na figura 1.5. A Equipe A é composta principalmente por cientistas de dados juniores, todos os quais ingressaram no mercado de trabalho sem um longo período acadêmico. Sua ação, ao obter os detalhes do projeto e as expectativas deles, é ir imediatamente para as postagens do blog. Eles pesquisam na Internet “detecção de fraude de pagamento” e “algoritmos de fraude”, encontrando centenas de resultados de empresas de consultoria, algumas postagens de blog de nível extremamente alto de cientistas de dados juniores semelhantes que provavelmente nunca colocaram um modelo em produção, e alguns rudimentares exemplos de dados de código aberto. A Equipe B, por outro lado, é composta por um grupo de pesquisadores acadêmicos com doutorado. A sua investigação e definição do âmbito são apresentadas na figura 1.6. Com a abordagem estudiosa da Equipe B para pesquisa e verificação de ideias, as primeiras ações são investigar artigos publicados sobre o tema da modelagem de fraude. Passando vários dias lendo jornais e artigos, esses membros da equipe estão agora munidos de uma grande coleção de teorias que abrange algumas das pesquisas mais avançadas realizadas na detecção de atividades fraudulentas. Se pedíssemos a qualquer equipe que estimasse o nível de esforço necessário para produzir uma solução, obteríamos respostas totalmente divergentes. A equipe A provavelmente estimaria cerca de duas semanas para construir seu modelo de classificação binária XGBoost, enquanto a equipe B contaria uma história muito diferente. Esses membros da equipe estimariam vários meses para implementar, treinar e avaliar a nova estrutura de aprendizado profundo que encontraram em um white paper altamente conceituado, cuja precisão comprovada para a pesquisa era significativamente melhor do que qualquer algoritmo implementado pelo Perforce para este caso de uso. O problema aqui com o escopo e a pesquisa é que esses dois polos opostos teriam seus projetos fracassados por dois motivos completamente diferentes. A equipe A falharia porque a solução para o problema é significativamente mais complexa do que o exemplo mostrado na postagem do blog (a questão do desequilíbrio de classe por si só é um tópico muito desafiador para ser documentado com eficácia no curto espaço de um blog). A Equipe B, embora sua solução provavelmente fosse extremamente precisa, nunca receberia recursos para construir a solução arriscada como um serviço inicial de detecção de fraudes na empresa.



Definir o escopo do projeto para ML é incrivelmente desafiador. Mesmo para os veteranos de ML mais experientes, conjecturar quanto tempo um projeto levará, qual abordagem terá mais sucesso e a quantidade de recursos necessários é um exercício fútil e frustrante. O risco associado a fazer afirmações errôneas é bastante alto, mas estruturar o escopo adequado e a pesquisa de soluções pode ajudar a minimizar as chances de se errar totalmente na estimativa. A maioria das empresas possui uma mistura de tipos de pessoas nesse cenário hiperbólico. Alguns são acadêmicos cujo único objetivo é promover o avanço do conhecimento e da pesquisa em algoritmos, abrindo caminho para futuras descobertas dentro da indústria. Outros são engenheiros de “aplicações de ML” que desejam apenas usar o ML como uma ferramenta para resolver um problema de negócios. É importante abraçar e equilibrar ambos os aspectos dessas filosofias em relação ao trabalho de ML, chegar a um acordo durante a fase de pesquisa e definição do escopo de um projeto e saber que o meio-termo aqui é o melhor caminho a ser trilhado para garantir que um projeto realmente dê certo para a produção.