

Trecho	Código	Categoria
sim, seguimos isso mesmo (sequência de etapas). De repente tem mais granularidade, mas de maneira geral é isso mesmo [...]. Por exemplo, o processamento de dados tem um conjunto de atividades [...], engenharia de atributos também, mas beleza.	Atribuiria mais granularidade ao workflow	Limitação [Geral]
não, não. Eu participava de todas as etapas mesmo. Desde o processamento dos dados, até a implantação do modelo. Às vezes a implantação não muito, mais o cara da infraestrutura. Então seria mais daqui (refere-se ao processamento de dados) até a avaliação do modelo. Em todos os projetos que eu vi, só tinha um cara pra fazer tudo isso.	Único desenvolvedor em todas as etapas.	Limitação [Geral]
	Presença do desenvolvedor	Ator [Geral]
	A implantação do modelo é auxiliada por uma equipe de infraestrutura.	Limitação [Implantação do modelo]
	Presença da equipe de infraestrutura	Ator [Implantação do modelo]
essas daqui: processamento de dados, engenharia de atributos, treinamento do modelo e avaliação. [...] Eu passava todos os pacotes para o cara (da infraestrutura) e ele implantava lá.	O modelo gerado é repassado para a equipe de infraestrutura para implantação	Atividade [Implantação do modelo]
eu participava da parte dos requisitos, porém mais para o entendimento. Ai tinham mais pessoas envolvidas no projeto, como o gerente	Desenvolvedor participa da elicitação de requisitos	Atividade [Requisitos do modelo]
	Gerente participa da elicitação dos requisitos	Atividade [Requisitos do modelo]
	Presença do gerente	Ator [Requisitos do modelo]
	Cliente participa da elicitação dos requisitos	Atividade [Requisitos do modelo]
	Presença do cliente	Ator [Requisitos do modelo]
cara, tipo assim, uma metodologia em si não. Era algo meio empírico, sabe?	Metodologias de desenvolvimento não são seguidas.	Atividade [Geral]
não, padrão de código não. [...] nós trabalhávamos com coisas mais ou menos conhecidas, como comentar o código [...], como era python toda essa questão da indexação do código tem que estar certinha [...], é mais isso.	Padrões de código não são seguidos.	Limitação [Geral]
Muitas vezes já aconteceu de nós estarmos treinando o modelo e a gente encontrar problemas, tendo que voltar para a engenharia de atributos. Isso também ocorreu na avaliação, pegar muita coisa que não está certa e voltar para a engenharia de atributos. E assim, é um vai e volta.	Ciclos de desenvolvimento existem devido a incerteza de correteude	Desafio [Geral]
	Problemas na avaliação indicam problemas na engenharia de atributos.	Limitação [Avaliação do modelo]
a avaliação. Se você não chega a um resultado bom, você tem que voltar, trocar modelo, ou mexer mais com os dados.	Avaliação do modelo como etapa que mais se repete.	Desafio [Avaliação do modelo]
no meu caso sim. Como a gente trabalhava com uma metodologia ágil, então temos o product owner. Então esse cara que fala se temos bons resultados ou não. É porque o cara conhece mais do negócio, entendeu? Ai o cara tá muito envolvido nos requisitos, e ele tem uma boa perspectiva do que o cliente quer.	Empresa desenvolve MLES.	
	O desenvolvimento do MLES segue uma metodologia ágil.	
	O cliente avalia a qualidade do modelo.	Atividade [Avaliação do modelo]
	Presença do cliente	Ator [Avaliação do modelo]
Por exemplo, a gente chegava a uma acurácia de 80%, mas pra o cara [...] isso não era suficiente. Porque o cara queria mais entender das causas do que a acurácia do modelo.	Métricas clássicas não suficientes para avaliar o modelo	Desafio [Avaliação do modelo]
	Interpretabilidade mais importante do que métricas.	Limitação [Avaliação do modelo]
cara, a engenharia de atributos. Porque deixar tudo certinho, fazer toda a limpeza, escolher as features importantes. Porque o treino do modelo é só rodar um algoritmo e pronto.	Processamento de dados como etapa mais problemática	Desafio [Processamento de dados]
	Engenharia de atributos como etapa mais problemática.	Desafio [Engenharia de atributos]
eu considero essa etapa mais problemática pelo esforço.	Esforço como palavra chave para definir etapa problemática.	Limitação [Geral]
na minha opinião é pegar um monte de atributos que não são relevantes para o modelo, que vão gerar muito ruído [...].	Problemas nos dados relacionados a ruídos	Desafio [Processamento de dados]
	Problemas nos dados relacionados a variáveis não representativas.	Desafio [Engenharia de atributos]

Então eu acho que isso é uma coisa que impacta demais no treinamento [...], na acurácia [...] e na avaliação.	Processamento de dados impacta nas demais etapas.	Limitação [Processamento de dados]
	Engenharia de atributos impacta nas demais etapas.	Limitação [Engenharia de atributos]
com certeza, aqui na engenharia de atributos. Se você tiver um monte de variáveis que fazem muito ruído é muito mais provável que teremos um overfitting no treino do modelo [...]. Nos olhamos overfitting também na árvore, na primeira versão que a gente rodou, a gente chegou em uma árvore muito desbalanceada, então isso [...] (indica) um overfitting muito grande, um ramo era muito grande [...].	Overfitting observado no balanceamento da árvore	Atividade [Avaliação do modelo]
	Origem do overfitting no processamento de dados.	Limitação [Processamento de dados]
nós fazemos análise de correlação e covariância e por conta disso a gente parte da ideia que ficou certo [...].	Uso de métodos estatísticos na engenharia de atributos	Método [Engenharia de atributos]
	Métodos estatísticos aplicados aos dados para evitar a existência de overfitting	Atividade [Engenharia de atributos]
Nós também fazemos uma validação cruzada com o cliente. [...] Porque nós não deixamos só que o algoritmo decida o que é ou não importante para o modelo. Já aconteceu do algoritmo falar para a gente que X variável não é importante, mas o cara lá do cliente fala, "essa variável é importante, você não pode tirar esse cara". Também acontece o contrário. Por exemplo, ele fala que algo (refere-se a um atributo) não é importante, mas [...] roda o modelo sem a variável e ela é importante.	Especialista do domínio indica atributos relevantes.	Atividade [Engenharia de atributos]
	Presença do especialista de domínio	Ator [Engenharia de atributos]
não, nesse caso não [...]. Só o Jupyter, Python.	Uso de frameworks e bibliotecas para a criação de modelos: Jupyter, Python	Métodos [Treino do modelo]
manipulando de novo né. Adicionando atributos, removendo atributos, normalizando os dados [...].	Problemas são corrigidos através de modificação e reexecução	Limitação [Geral]
isso, a gente chegou nessa conclusão. Se o cara quer entender das causas [...], nós não podíamos usar uma rede neural, porque nós chegaríamos a um resultado, mas a interpretação desse cara aí seria complicada [...]. Então por conta disso, a gente escolheu árvore de decisão.	Árvore de decisão necessária devido ao requisito de interpretabilidade	Limitação [Geral]
	Redes Neurais não provêem interpretabilidade.	Limitação [Geral]
nesse caso o product owner que fala se está bom ou não, de acordo com o que ele conhece do cliente.	O cliente define os melhores valores para as métricas de avaliação.	Atividade [Avaliação do modelo]
então, nessa primeira entrega, nós chegamos a um modelo não tão bom [...].	Modelo não satisfatório em termos de regressão, utilizando árvore de decisão	Limitação [Geral]
Como eu estava falando, naquela questão das causas, nós temos um modelo que é bom, ele explica bem o que está acontecendo, mas a previsão no modelo de regressão é ruim. Nesse caso nós entregamos essa v zero que atende a explicação das causas, mas a predição e a regressão não foi boa. Então estamos vendo se mudamos de algoritmo, como pegar uma rede neural [...].	Modelo com boa interpretabilidade	Limitação [Avaliação do modelo]
	Modelo com baixa precisão de regressão	Limitação [Avaliação do modelo]
	Rede Neural como Alternativa para regressão.	Limitação [Avaliação do modelo]
Na próximas etapas, vamos ver se exploramos mais a engenharia de atributos, dependendo do que nós encontramos ali, a gente continua, ou muda de algoritmo [...].	Incerteza sobre a qualidade da engenharia de atributos.	Desafio [Engenharia de atributos]
árvore de decisão é muito mais simples de interpretar do que outros algoritmos, mesmo random forest tendo estrutura de árvore ele cria várias, então a interpretação não é tão simples assim, como tendo só uma árvore [...].	Árvore de decisão como a melhor alternativa para interpretabilidade.	Limitação [Requisitos do modelo]