

Trecho	Código	Categoria
é... vamos lá né... Nesse workflow, a gente tá falando mais especificamente de ML, né? Então, no final das contas, é... esse workflow representa um subconjunto dos nossos workflows, a gente tem uma equipe ali desenvolvimento de software que não faz somente ML, então acho que dentro desse subconjunto.	Empresa desenvolve MLES.	
	Modelos de ML são módulos do MLES.	
Começando ali pelo primeiro, a parte de requisitos [Model Requirements], eu acho que vem muito do cliente, né? O que o cliente quer, qual a tarefa que ele gostaria que fosse atacada usando ML, qual é o problema que ele está enfrentando e que a gente definiu internamente conversando que ML é a melhor ferramenta para atacar aquele problema e etc, né? Acho que tem muito do cliente, do input né? Do... do... de qual é o problema, o que é que a gente quer melhorar e etc, né? E aí eu acho que com base nisso, já... já entra ali o pessoal né?	Requisitos do modelo como a definição das tarefas que devem ser resolvidas com ML.	
	Requisitos do modelo especificados pelo cliente.	Atividade [Requisitos do modelo]
	Presença do cliente	Ator [Requisitos do modelo]
	Presença do desenvolvedor	Ator [Requisitos do modelo]
A gente não tem ali papéis bem definidos, como uma software house teria, por ser um grupo que você sabe, né? Acho que os outros que você entrevistou também são do grupo ali, é um grupo de pesquisa aplicada, né? Tem muito desse componente de computação mas também tem um outro componente de pesquisa, não é uma software house por assim dizer, né? Então esses papéis se confundem um pouco.	Empresa de pesquisa e desenvolvimento	
	Desenvolvedores não possuem papeis bem específicos relacionados a ML	Limitação [Geral]
A gente tem... pessoas ali - que vamos dizer assim - tem esse interesse, essa área de atuação mais voltada para ML e qualquer uma dessas pessoas - respondendo a sua pergunta - atuam na parte de entender os requisitos e pensar quais modelos de ML são os mais adequados e etc, né? Então qualquer uma dessas pessoas, seja com um nível de mestrado ou doutorado podem vir a atacar esse problema e começar ali a estudar como propor um modelo de ML, para resolver o problema apresentado pelo cliente.	Desenvolvedores com interesse em ML trabalham na criação dos modelos.	Limitação [Geral]
Eu falei muito dos requisitos, né? Mas os outros são a mesma coisa, a gente tem esse grupo de pessoas, que tem essa - vou dizer aqui, interesse, área de atuação, que faz pesquisa - que sabe o que é ML, que entende como funciona essa tecnologia e tal. Sendo que elas atacam muitas vezes esse problema, muitas vezes pode ser um grupo de pessoas que sentam para discutir, às vezes uma única pessoas.	Único desenvolvedor participa de todo o workflow como alternativa	Limitação [Geral]
	Vários desenvolvedores participam de todo o worflow como alternativa.	Limitação [Geral]
A coleta de dados liga muito ao cliente, né? A gente volta para o cliente e pede, né? "Ah, a gente precisa de um dado assim, de um dado assado, com tal característica para testarmos tal hipótese e etc". Muitas vezes a rotulação dos dados ela é feita pelo cliente, acho que é uma característica também que vale a pena mencionar, porque é ... se você fala de aplicações de ML mais gerais, que você tem, vamos lá, aqueles exemplos bem clássicos da imagem do gato, imagem do cachorro o carro e etc. Qualquer pessoa pode rotular, mas se você tem um negócio mais de domínio, que tem uma expertise específica, você vai precisar que o cliente ou especialista do domínio te ajude a rotular, nós mesmos não temos a capacidade de rotular.	Processamento de dados feito pelo cliente	Atividade [Processamento de dados]
	Fornecimento dos dados feito pelo cliente	Atividade [Processamento de dados]
	Cliente/Especialista do domínio rotula os dados em domínios específicos.	Atividade [Processamento de dados]
	Presença do cliente	Ator [Processamento de dados]
	Presença do especialista de domínio	Ator [Engenharia de domínio]
A parte do model training e model evaluation é a mesma coisa que eu falei da parte de requisitos. E a feature engineering, também acho que vale a pena ressaltar, que hoje em dia a gente fala muito em deep learning e quando você fala de deep learning você pula um pouco essa etapa, né? Você tem ali um modelo que aprende as features, e também aprende a tarefa, seja ela uma tarefa de classificação, seja ela uma tarefa de ... segmentação, regressão e etc. A gente nem sempre passa por essa etapa de <u>Feature Engineering, ela pode ser pulada</u> , como eu te falei, em geral, se eu pego um problema de ML para tratar, eu lido com todas essas etapas.	Único desenvolvedor participa de todo o workflow como alternativa	Limitação [Geral]
	Vários desenvolvedores participam de todo o worflow como alternativa.	Limitação [Geral]
	Deep learning diminui a necessidade de engenharia de atributos	Limitação [Engenharia de atributos]
	Modelos de deep learning aprende as features e as tarefas	Limitação [Treino do modelo]
	Deep learning para engeharia de dados e treino do modelo de maneira única.	Método [Geral]
	Uso de deep learning para criação de modelos	Atividade [Treino do modelo]
	Único desenvolvedor em todas as etapas do workflow como alternativa.	Apontamento [Desenvolvimento]

<p>A &lt;&lt;compahia&gt;&gt; é uma empresa grande, né? E a gente tem uma infraestrutura em nuvem, tem toda essa complexidade para você fazer o deployment de um modelo. A gente tem pessoas lá no grupo que... mais técnicas, que ajudam a gente, entende como é que essa infraestrutura funciona, como é que faz para colocar um modelo disponível em um serviço na nuvem e etc. Nessa etapa tem um pessoal mais especializado, mas até a parte de avaliação... desde o requisito até a parte de avaliação é a mesma pessoa, ou grupo de pessoas, caso seja um problema que tem mais de uma pessoa tratando dele.</p>	Equipe especializada implanta modelo em núvem	Limitação [Implantação do modelo]
	Presença da equipe de infraestrutura	Ator [Implantação do modelo]
<p>dessas etapas, eu acho que... vamos lá, deixa eu pensar um pouquinho aqui... Eu acho que a maioria das pessoas vai te falar que Data Cleaning é um dos principais problemas de ML e eu concordo, tá?</p>	Processamento de dados como etapa mais problemática.	Desafio [Processamento de dados]
<p>Porém como eu falei, os problemas que - eu pelo menos - tive a oportunidade de trabalhar na &lt;&lt;compahia&gt;&gt; a gente recebe muita coisa já pronta, ou trabalhada pelo cliente, então eu não vejo essa etapa - pelo menos do ponto de vista de ML - como uma das que mais tomam tempo.</p>	Cliente executando o processamento de dados, diminui a complexidade da tarefa	Limitação [Processamento de dados]
<p>Então eu acho que eu escolheria aí o Feature Engineering, eu passei bastante tempo pesquisando features e trabalhando com features novas, combinações de features e avaliando quais delas iriam produzir a melhor acurácia, então seria a caixinha aí de Feature Engineering.</p>	Entrevistado considera ter mais experiência na Engenharia de atributos.	Apontamento [Desenvolvimento]
<p>a gente... bem, depende de como você vê o Feature Engineering, né? Acho que um extremo seria você propor a sua própria feature, né? Propor uma forma de extrair uma característica do dado que vai trazer para você uma possibilidade de discriminação maior, né?</p>	Atributos são aprendidos nos dados	Atividade [Engenharia de atributos]
	Atributos são coletados dos dados	Atividade [Engenharia de atributos]
<p>A gente não chega nesse ponto, então a gente faz muito, é... pegar um conjunto de features... tem uma etapa de pesquisa, né? Quais são as features que a literatura aponta como boas para essa tarefa, dessas features que a literatura aponta a gente faz experimentos com elas individualmente avaliando a acurácia, e muitas vezes também combinando, né?</p>	Busca na literatura de atributos conhecidos como relevantes para o problema.	Atividade [Engenharia de atributos]
<p>Para ver se combinando a gente também produz uma acurácia maior. Uma outra possibilidade é a gente usar um seletor de features - por assim dizer - que aponta ali as features que tem a maior importância para aquela tarefa, então geralmente essa etapa caminha muito nessa direção de avaliar e combinar essas features e selecionar as melhores para aquela tarefa.</p>	Uso de seletor de atributos para identificar atributos mais relevantes.	Atividade [Engenharia de atributos]
	Seletor de features para a escolha de atributos.	Método [Engenharia de atributos]
<p>é... tem duas formas que eu consigo pensar rapidamente. A primeira - acho que é a mais fácil - é aquela que eu já falei que você tem algum tipo de ground truth que a gente chama de padrão ouro, ou verdades de campo e a gente utiliza essa informação como entrada para um seletor de features, e esse seletor de features ele já dá para você, com base nessa verdade de campo quais são as features que parecem ser mais importante para aquela determinada tarefa, por exemplo, um problema de classificação, e aí a gente caminha nessa direção.</p>	Seletor de atributos aprimorados com verdades de campo	Método [Engenharia de atributos]
<p>Então vê se eu enderecei direito a sua pergunta. Então esse cara, essa resposta desse seletor de features já é uma forma de avaliação, tá? Da importância daquela features.</p>	Seletor de atributos aprimorado com verdades de campo como verificador dos melhores atributos para treinar o modelo.	Atividade [Engenharia de atributos]

Uma outra forma é, eu abro mão do seletor de features e faço na mão então eu vou selecionando ali subconjuntos de todas as features que eu quero avaliar, treino o modelo e na avaliação do modelo eu avalio se aquele conjunto de features produziu uma acurácia maior ou menor do que os outros conjuntos testados	Seleção manual de atributos	Atividade [Engenharia de atributos]
No final das contas é meio que uma força bruta, né? Você vai chegar em um ponto que você vai tem ali uma ideia de quais foram as features que produziram os melhores modelos.	Seleção manual de atributos acontece por força bruta	Limitação [Engenharia de atributos]
eu acho que... bom tem alguns, né? O primeiro acho que é essa pesquisa que tem que ser feita, se você for do domínio... se você já é do domínio... eu trabalho muito com análise de imagens, e já tenho uma boa noção que funcionam bem para determinadas tarefas, né? Mas se você tá pegando um tipo de dado novo, por exemplo, vou trabalhar com áudio, vou trabalhar com processamento de texto, com vídeo. Factualmente eu já não vou ter essa ideia, né? De quais são as melhores features, então vai ter uma etapa de pesquisa, geralmente indo a livros a papers e etc, para você fazer essa seleção, essa pré seleção de features. Então acho que esse é o primeiro desafio se você não é do domínio, se você é do domínio, você já tem uma boa ideia.	Pertencer ou não ao domínio do projeto é um primeiro desafio enfrentado pelo desenvolvedor.	Desafio [Engenharia de atributos]
	Necessidade de estudar novos temas para lidar com alguns projetos.	Desafio [Engenharia de atributos]
O outro desafio é fazer esses experimentos, muitas vezes - como eu falei - é força bruta mesmo, você ir testando várias possibilidades, é... você nunca sabe qual é a melhor feature para aquela tarefa, a não ser que você teste exatamente para aquela tarefa.	Identificar as melhores atributos para treinar o modelo como um desafio	Desafio [Engenharia de atributos]
Você pode ter feito um conjunto de features que foi muito bem em uma outra tarefa até similar a ela, mas uma pequena alteração na característica do dado ou na tarefa que você quer executar, pode é... representar algum tipo de mudança no conjunto de features que vai gerar o melhor resultado naquele momento, então acho que é essa...	O mesmo conjunto de atributos podem não funcionar para problemas similares	Limitação [Engenharia de atributos]
	Características dos dados influencias nos atributos.	Limitação [Engenharia de atributos]
Então o primeiro desafio acho que é você descobrir quais são as features, o outro é desse subconjunto que você descobriu que atendem aquele tipo de dado, qual o que de fato vai ter os melhores resultados para os seus dados.	Encontrar um subconjunto dentro dos atributos relevantes que realmente são os melhores para o problema.	Desafio [Engenharia de atributos]
Uma outra questão é que geralmente é um processo lento, em última instância você está reduzindo a dimensionalidade do seu dado, vamos trabalhar com imagens ou vídeo por exemplo, você tem um dado n-dimensional, duas dimensões, três dimensões, dependendo e você quer reduzir para um único descritorzinho que vai ser potencialmente discriminate para você separar as classes de interesse, né? Então esse processo de computação, em termos de performance computacional também é um processo caro, a gente tenta paralelizar e etc para reduzir um pouco esse esforço.	A transformação de atributos é um processo caro computacionalmente.	Desafio [Engenharia de atributos]
	Uso de transformação de atributos.	Método [Engenharia de atributos]
	Tranformação de atributos para reduzir os atributos	Atividade [Engenharia de atributos]

<p>tá... então... acho que o processo é o mesmo que a gente discutiu, se você tá usando um seletor de features por exemplo, ele vai... você pode avaliar a relevância de uma feature de acordo com a importância que aquele seletor te dá, geralmente isso é apresentado em forma de uma nota, de um score, né? Você tem ali o maior score, aquela feature é mais importante, mais relevante, menor score, aquela feature é menos relevante, então usando algum desses métodos, você já tem diretamente aquela medida de relevância.</p>	<p>O seletor de atributos indica atributos relevantes e irrelevantes</p>	<p>Limitação [Engenharia de atributos]</p>
<p>No ponto em que você faz isso mais manualmente, como eu estava falando, fica muito a critério de como você faz o design, o planejamento do seu experimento, né? Mas você vai fazendo ali combinações de features e vai avaliando a acurácia, a qualidade do resultado usando aquelas features específicas e depois de alguns experimentos você vai ver que tem um conjunto de features que quando elas são selecionadas para serem utilizadas, a acurácia é menor, né? Então você tem ali um indicativo que aquelas features são menos relevantes, ou menos discriminantes para aquele problema específico.</p>	<p>Durante a seleção manual de atributos a relevância é indicada através de treino e avaliação.</p>	<p>Limitação [Engenharia de atributos]</p>
<p>eu acho que geralmente a gente vai fazendo uma coisa que meio, por demanda né? Você faz o primeiro teste, se você ver que o resultado ainda não é bom, você vai atacando os possíveis problemas.</p>	<p>O processo de melhoria dos atributos é iterativo.</p>	<p>Limitação [Engenharia de atributos]</p>
<p>Um deles é a normalização das features, né? As vezes você tem que fazer algum tipo de normalização, entre zero e um, ou aquela padronização do dado onde você subtrai a média e divide pelo desvio padrão. Dependendo da tarefa um tipo de normalização é melhor que outro, ou até mesmo tem normalizações que pioram o resultado, então geralmente é uma questão de tentativa e erro. A gente faz muito isso de acordo com o que os experimentos vão dizendo para a gente, agora existe também uma questão que eu acho importante nesse ponto que você levantou, que é...</p>	<p>Normalizações utilizadas para ajustar os atributos</p>	<p>Atividade [Engenharia de atributos]</p>
	<p>Uso de técnicas de normalização dos atributos</p>	<p>Método [Engenharia de atributos]</p>
<p>.. o conjunto de features muitas vezes ele é grande, não dá para a gente é... sair usando todas as features que a gente tem disponível para aquele problema, então muitas vezes a gente usa um redutor de dimensionalidade, como a gente diz, com o PCA por exemplo, que ele já faz um pouco dessa limpeza dos dados. Ele vai deixar... ele aplica uma transformação nas features que projeta essas features em um espaço de dimensão menor do que o espaço original, onde essas features são as mais relevantes, ou melhor, explicam melhor a variância apresentada nos dados. Então com isso em teoria você já está tirando features que são irrelevantes, né? Que são correlacionadas.</p>	<p>Dados com alta quantidade de atributos são transformados através de redutores de dimensionalidade.</p>	<p>Atividade [Engenharia de atributos]</p>
	<p>Uso de PCA para transformação nos atributos.</p>	<p>Método [Engenharia de atributos]</p>
	<p>Redutores de dimensionalidade removem dependência nos dados.</p>	<p>Limitação [Engenharia de atributos]</p>
<p>essa é uma parte... eu sei que isso existe, na minha educação de Machine Learning eu trabalhei com isso, mas na prática eu não... eu particularmente não usei muito. Tipo, pegar uma feature e transformá-la.</p>	<p>Transformação de atributos mais complexas não são aplicadas</p>	<p>Limitação [Engenharia de atributos]</p>
<p>Agora uma coisa que a gente faz, claro, você trabalha com dados nulos, então você tem dados faltantes, você tem que aplicar algumas transformações para que você possa garantir que você tenha dados válidos. Uma das transformações que são aplicadas é... sei lá, usar o valor médio daquela feature ou a mediana e por aí vai. Tem algumas funções possíveis,</p>	<p>Transformações aplicadas em dados nulos ou faltantes</p>	<p>Atividade [Engenharia de atributos]</p>

<p>mas uma vez que você tem todos os valores válidos ali é... a gente não trabalha muito em cima das features, talvez porque eu trabalhe muito com imagem, então na verdade essa... essa brincadeira com transformações ela acaba acontecendo antes - muitas vezes - das features serem extraídas. A gente aplica na própria imagem, a gente aplica um realce na imagem, a gente aplica uma transformação na imagem que aumenta o contraste da imagem, entendeu? Então a gente tem algumas transformações no dado em si e não necessariamente nas features que são extraídas daqueles dados.</p>	Em imagens transformações são feitas nas próprias, não em atributos extraídos.	Atividade [Engenharia de atributos]
	Enriquecimento de dados.	Método [Processamento de dados]
<p>essa pergunta ela é interessante, porque eu acho que é um profissional que está com os dias contados. No sentido de que eu falei anteriormente do deep learning por exemplo, né? Para muitas aplicações, não todas, por isso que eu falei que é uma resposta difícil de dar, mas a tendência é que esse profissional ele seja cada vez menos requisitados, porque nós temos algoritmos que conseguem não somente aprender a resolver a tarefa - como eu tava comentando - mas também conseguem aprender quais são as melhores features para resolver aquela tarefa, então você meio que abriu mão completamente desse profissional.</p>	Especialistas de domínio são menos necessários em modelos de deep learning	Limitação [Engenharia de dados]
	Presença do Especialista do domínio	Ator [Engenharia de dados]
	Deep learning aprende as features e a resolução da tarefa.	Método [Engenharia de atributos]
<p>Por outro lado, a gente sabe que deep learning e essas extração automática de features, ela não resolve o problemas do mundo, então ainda há espaço para esse profissional,</p>	Deep learning não resolve todos os problemas, isso faz necessário um especialista de domínio.	Limitação [Engenharia de atributos]
<p>mas acho que cada vez em nichos mais específicos, em problemas mais específicos, principalmente aqueles problemas de domínio que eu comecei falando, em que você ter uma expertise que não é facilmente aprendida, né? Por um profissional de TI, por um data scientist, por um machine learning engineer e etc, eu acho que esse profissional que entende os dados, né? Que entende as limitações do domínio, vamos supor por exemplo que é um domínio que também além dos dados em si claro, né? Você tem leis físicas que regem aquele domínio de dados ali e que essas leis físicas precisam ser respeitadas nessas features, né? Então seria acho que uma área que esse profissional seria requisitado ou até mesmo necessário.</p>	Domínio específicos fazem necessário o especialista de domínio.	Limitação [Engenharia de atributos]
<p>sim, a gente trabalha em geral com o domínio de recursos naturais no grupo, né? E a gente trabalha em geral com geólogos e geofísicos.</p>	Geólogos como especialistas de domínio	Ator [Engenharia de atributos]
	Geofísico como especialista de domínio	Ator [Engenharia de atributos]
<p>Esses geólogos e geofísicos trazem o conhecimento da geologia da geofísica, e eles muitas vezes, eles nos alertam, "olha uma solução data driven, simplesmente baseada em dados nesse caso aí é um pouco problemático porque a gente tem toda uma física que explica esses dados que precisa ser respeitada".</p>	Questões do domínio precisam ser respeitadas pelo modelo	Limitação [Treino do modelo]
	Soluções totalmente baseadas em dados podem não ser ideais	Limitação [Treino do modelo]
	Especialista nos dados indicam regras que precisam ser respeitadas	Atividade [Treino do modelo]
	Soluções data-driven nem sempre são possíveis.	
<p>Até hoje tem um movimento aí que o nosso grupo também trabalha - talvez fosse até legal você entrevistar alguém que tá vendo isso, o &lt;&lt;cientista de dados&gt;&gt; é uma referência, mas a &lt;&lt;cientista de dados&gt;&gt; pode te dar o ponteiro para ele - que está trabalhando com modelos de ML physically informed, né? Então você usa a física, leis físicas para informar e reger por assim dizer o modelo de ML, para que ele sempre respeite essas leis físicas, então já tem uma linha de pesquisa voltada para isso.</p>	Modelos Physically informed são regidos por regras físicas.	

<p>é... como eu falei, acho que... pelo menos eu na minha experiência, tem isso né? A gente pode usar um modelo que já aprende as features, a gente pula completamente essa etapa. Tem a possibilidade de pegar as features e usar um seletor de features, como a gente chama, né? Que vai te dizer ali quais são as features mais relevantes para aquele problema e tem a força bruta como eu falei, que é você extrair as features, treinar o modelo, avaliar o modelo e baseado naquilo avaliação você determina se aquela feature foi boa ou não. Acho que é mais ou menos isso.</p>	Incerteza sobre se os atributos escolhidos são os melhores para o modelo final	Desafio [Engenha de atributos]
<p>essa é uma pergunta interessante, porque... talvez por eu não ser um engenheiro de software por assim dizer de profissão, minha posição é de pesquisador e o software acaba sendo uma ferramenta, a programação acaba sendo uma ferramenta para alcançar os resultados de pesquisa desejados, né?</p>	Entrevistado não é da área de computação	Apontamento [Engenharia de atributos]
	Uso de ML como ferramenta para solução de problemas	
<p>Então a gente acaba se baseando muito em bibliotecas disponíveis, né? Como scikit-learn, até na parte de imagem tem o scikit-image, o numpy, o scipy e etc, então assim eu não me vejo muito olhando para esses códigos diretamente, né? Eu programando esses códigos, então geralmente eu leio sobre o que o método faz, entendo o que o método faz ou lendo a própria documentação dessas bibliotecas, ou lendo os artigos que deram origem aos métodos e uma vez que eu entendi, eu utilizo se já estiver disponível. Acaba que programar essas coisas na mão fica cada vez menos comum para mim pelo menos.</p>	Uso de frameworks para a criação de modelos.	Método [Geral]
	Incerteza sobre a correteza da implementação do framework.	Desafio [Treino do modelo]
então, na nossa experiência isso vai muito do cliente, como eu falei.	Primeiro responsável pelo processamento de dados é o cliente.	Limitação [Processamento de dados]
Os dados que a gente coleta estão intimamente relacionados ao problema que o cliente apresenta, não é? Baseado neste problema a gente verifica quais são os dados que melhor iriam ajudar a gente a ter uma primeira ideia de qual é o problema, como atacar o problema, as melhores features, modelos e etc.	A primeira análise dos dados é o que define como ele será abordado	Limitação [Processamento de dados]
Acho que aí é uma colaboração muito próxima com o cliente, que na verdade é o expert naquele domínio e conhece o dado melhor do que ninguém.	Cliente do domínio é quem melhor entende o dado.	Limitação [Processamento de dados]
<p>Na parte de limpeza e labeling, como eu falei, ela acaba vindo muito do cliente, não é? A gente tem exemplos em que a gente recebeu dados e a gente rodou um modelo de ML, alguma coisa e percebeu que o resultado não estava interessante, voltou para o cliente e ele falou "ah, realmente gerei o dado de maneira errada ou esqueci de avisar tal coisa".</p>	O cliente executa as etapas de coleta, limpeza e rotulação dos dados	Atividade [Processamento de dados]
	Cliente pode cometer erros no processamento de dados.	Limitação [Processamento de dados]
Na nossa vivência - no laboratório com os clientes que nós temos tido - isso acaba ficando por conta deles,	É característica dos projetos sempre receber os dados processados do cliente.	Limitação [Processamento de dados]
essa etapa, porém eu mencionei anteriormente que muitos que trabalha com ML veem essa etapa de limpeza, coleta de dados e rotulação como a mais trabalhosa e mais importante, porque nem sempre você tem o cliente ou o especialista no dado a disposição para você.	Processamento de dados como etapa problemática se você não tem auxílio do cliente.	Desafio [Processamento de dados]
Você precisa fazer isso manualmente, então... mas assim, que eu tente puxar aqui pela memória rapidamente, a gente não trabalhou muito nessa direção de limpeza, coleta e rotulação de dados, a gente diretamente.	Na empresa pouco se trabalhou sem o cliente no processamento de dados.	Limitação [Processamento de dados]

é, essa pergunta é interessante também, no sentido de que se você já é do domínio - é a segunda vez que eu volto a esse ponto - se você conhece qual é a característica esperada daquele dado, se você já trabalhou em algo parecido antes, tudo fica mais fácil.	Ter experiência no domínio é de grande utilidade no processamento de dados	Limitação [Processamento de dados]
Por exemplo, hoje a gente tem trabalhado com imagens sísmicas, ao olhar para uma imagem hoje eu já sei se ela tem a qualidade necessária para gerar um bom resultado, se eu vou conseguir tirar um bom resultado dela ou se não, se eu vou ter que voltar para o cliente e dizer, "olha não tem como processar esse dado? Não tem como entregar um dado com maior qualidade?".	Entrevistado diz ter experiência na área que atua.	
	Desenvolvedor consegue perceber a qualidade de imagens úteis para a área que ele conhece	Limitação [Processamento de dados]
	Melhorias de qualidade nos dados são solicitadas ao cliente.	Atividade [Processamento de dados]
Agora se você não é do domínio essa avaliação é muito difícil, você não sabe qual é a característica de um dado bom ou um dado ruim, então nesse caso você não tem muito o que fazer.	Desenvolvedores que não são do domínio dificilmente percebem a qualidade nos dados que tem.	Limitação [Processamento de dados]
Acho que um caminho possível é tentar rodar o modelo, rodar todo o workflow, essa pipeline aí de aprendizagem de máquina e ver se consegue um bom resultado, se você conseguir um bom resultado talvez isso indique que o dado é adequado. Se você não conseguir um bom resultado, acho que é o momento de voltar para quem te gerou o dado, né? Que geralmente no nosso caso é o cliente, e perguntar se não tem como... se ele não tem algum problema no dado ou se ele não consegue gerar um dado com melhor qualidade. Nesse caso aí é tentativa e erro, não tem jeito.	Desenvolvedores que não são do domínio tem que avaliar a qualidade do dados executando o workflow.	Atividade [Processamento de dados]
eu acho que é isso, né? Você criar essa... essa experiência com o dado a ponto de você começar a ter um sentimento do que é um dado com boa qualidade, o que é um dado com baixa qualidade, é... qual é o resultado esperado.	Um desafio do processamento de dados é adquirir essa capacidade de perceber a qualidade do dado de certo domínio.	Desafio [Processamento de dados]
Por exemplo, tem problemas de ML que são tão complicados que mesmo com dados de boa qualidade, você não consegue uma acurácia tão alta, porque o problema é complicado, então uma vez que você já tenha essa noção a priori, você já sabe o que esperar.	Alguns problemas de ML são complicados por natureza e nem dados de boa qualidade rendem bons resultados	Limitação [Geral]
Se você não tem, e é a primeira vez que você está trabalhando com aquele dado, se você tem uma acurácia baixa já acende uma alerta de que o dado pode estar com problema, ou tem um problema no seu modelo. Mas na verdade não é, é o melhor que se consegue porque o problema é complicado, né? Acho que a experiência da pessoa trabalhando naquele dado, naquele projeto, com aquele tipo de problema é fundamental para que você consiga enxergar essas coisas de maneira mais fácil.	Trabalhar com problemas complicados pode gerar a falsa impressão que algo de errado existe no workflow.	Limitação [Geral]
Caso você não tenha essa experiência, é tentativa e erro mesmo, não tem jeito. Ou claro, você pode acionar um colega, caso você tenha alguém disponível que tenha essa experiência, ou até mesmo, como eu já falei voltar ao cliente, tentar entender um pouco melhor, ou até mesmo para artigos e livros que possam ajudar nesse processo.	Se não há experiência no domínio o trabalho é de tentativa e erro	Limitação [Geral]
	Artigos e livros são úteis para aprender sobre um domínio.	