

Trecho	Código	Categoria
Nessa engenharia de atributos é que você tem toda essa parte de estatística, conhecimento dos seus dados, redução dos dados [...].	Engenharia de atributos como a etapa de análise dos dados.	Atividade [Engenharia de atributos]
O que é importante para Machine Learning? Grande volume de dados, se você não tiver, você não pode utilizar [...].	Volume de dados como aspecto crucial para uso de ML.	Limitação [Geral]
Em uma área exploratória não funciona (refere-se a Machine Learning) porque eu não tenho dados, entendeu? Eu consigo isso se eu tenho muitos dados na área.	Áreas explorativas dependem da quantidade dos dados.	Limitação [Geral]
bom, no nosso caso nós conseguimos uma forma de programação que a gente pode distribuir (desenvolvedores) para cada uma (refere-se às etapas) [...], mas todo mundo sabe entrar em cada módulo desse.	Desenvolvedores diferentes para cada uma das etapas como alternativa	Limitação [Geral]
	Presença do desenvolvedor	Ator [Geral]
	Cada desenvolvedor conhece todas as etapas.	Limitação [Geral]
Nós fazemos nessa sequência, mas aí um pode ficar na estatística, outro pode ficar na seleção, outro pode fazer os modelos lá de classificação, mas [...] eles podem compartilhar com todo mundo.	Desenvolvedores diferentes para cada uma das etapas	Limitação [Geral]
	Cada desenvolvedor deve se comunicar com todos os demais.	Limitação [Geral]
sim, no nosso caso como nós trabalhamos com a <<compahia>>, tem alguém de lá que vai analisar a gente roda o modelo discute com eles e então eles vão analisar tudo.	Cliente analisa a qualidade do modelo gerado.	Atividade [Avaliação do modelo]
	Presença do cliente	Ator [Avaliação do modelo]
Eles vão dizer: olhe podemos utilizar essa variável, podemos retirar essa variável, pela experiência deles.	Modelo posto em execução para análise do cliente.	Atividade [Avaliação do modelo]
De acordo com sua experiência o cliente pode indicar melhorias em relação aos atributos do modelo.		Atividade [Avaliação do modelo]
bem, nessa área hoje estamos Eu e o <<cientista de dados>>, a gente discute [...].	Internamente modelo avaliado por todos os desenvolvedores.	Atividade [Avaliação do modelo]
sim, as etapas se repetem. Você volta, muda as variáveis ou o modelo que pode usar, e testar esse conjunto de dados. [...]	Ciclos de desenvolvimento existem devido a incerteza de correteude.	Desafio [Geral]
Posso fazer uma outra redução, um PCA, um discriminate [...]. Nessa parte isso é bem cíclico.	Uso de métodos estatísticos na Engenharia de Atributos.	Atividade [Engenharia de atributos]
Isso aqui ó, [...] aqui sim você fica trabalhando: nos atributos, treino e avaliação do modelo, nestes três a gente fica trabalhando aqui.	Avaliação da indícios da correteude do modelo	Limitação [Avaliação do modelo]
[...] (devido a) quantidade de dados, qualidade dos dados, redundância dos dados, e o que a gente chama de [...] correlações entre os dados [...].	Quantidade dos dados como aspecto importante	Limitação [Processamento de dados]
	Qualidade dos dados como aspecto importante	Limitação [Processamento de dados]
	Correlação entre os dados como aspecto importante.	Limitação [Engenharia de atributos]
Porque a gente pensa que pega os dados seleciona os atributos e joga, gerou o modelo [...]. Se você achou que tá um modelo bom pra esse conjunto de dados, você vai aplicar. Depois quando você vai aplicar em outros dados, você ver que seu modelo não está ajustado, porque você não tratou a casualidade, redução dos dados [...].	Overfitting causado pela falta de cuidado no tratamento dos dados.	Desafio [Engenharia de atributos]
Eu acho que essa é a parte mais importante e mais trabalhosa, processamento de dados e seleção de atributos	Processamento de dados é a etapa mais trabalhosa	Desafio [Processamento de dados]
	Engenharia de atributos é a etapa mais trabalhosa.	Desafio [Engenharia de atributos]
Porque o resto você fica mudando parâmetro, usa essa rede, usa essa.	Definição empírica de hiperparâmetros para os modelos	Atividade [Treino do modelo]
	Definição empírica de arquiteturas para os modelos	Atividade [Treino do modelo]
ruídos, [...] falta de dados, [...] erros nos dados, muitos erros. Variáveis que não tem causalidade, que não tem relação nenhuma.	Problemas nos dados relacionados a erros	Desafio [Processamento de dados]
	Problemas nos dados relacionados a ruídos	Desafio [Processamento de dados]
	Problemas nos dados relacionados a variáveis pouco representativas.	Desafio [Engenharia de atributos]
estatística multivariada, PCA, análise de discriminantes. Tudo da estatística multivariada, [...]	Métodos estatísticos utilizados para identificar problemas nos dados.	Método [Engenharia de atributos]
eu conhecendo o meu dado eu posso identificar os melhores atributos que eu vou utilizar.	Conhecer o dado auxilia a identificar os melhores atributos.	Limitação [Engenharia de atributos]
gente usa precisão, recall, analisa a matriz de confusão, gera vários modelos e usa a Curva ROC.	Métricas clássicas são utilizadas na avaliação do modelo	Atividade [Avaliação do modelo]
	Uso de métricas clássicas na avaliação.	Método [Avaliação do modelo]
a gente utiliza tudo que tem no Python, tudo Python aqui. [...] O Python já nos disponibiliza tudo.	Uso de python e frameworks para a criação do modelo.	Método [Treino do modelo]
Eu uso o Python, como eu tô com meus dados no Pandas eu consigo usar só as minhas variáveis (refere-se às variáveis de interesse do modelo) [...], eu discrimino, eu nomeio as variáveis de interesse.	Alteração de dados para corrigir problemas.	Atividade [Engenharia de atributos]
com os dois, nós trabalhamos com classificação, e muita coisa é regressão.	Uso de modelos para atividades de classificação e regressão.	Método [Geral]