



[Return to "Machine Learning Engineer Nanodegree" in the classroom](#)

# Machine Learning Capstone Project

## REVISÃO

## REVISÃO DE CÓDIGO

## HISTORY

### Meets Specifications

Olá Lucas,

Parabéns! Você fez as últimas alterações e concluiu com sucesso o projeto final! Além disso, eu pessoalmente gostei muito de ler o seu relatório. Adicionei sugestões para projetos futuros ao longo dessa revisão, espero que elas lhe sejam úteis. Por fim, desejo todo o sucesso na sua carreira :)

Att,

[Heitor](#)

slack: hmgomes

### Definição do problema

O relatório contém uma explicação em alto nível do projeto em termos leigos, com informações sobre o domínio do problema, sua origem e o conjunto de dados utilizado.

### Excelente

- Você explica desde a origem do aumento do problema do fluxo migratório até a parte de controle de visto. Ótimo trabalho!
- As Fig 1 e 2 foram ótimas adições a discussão.

O problema abordado é definido com clareza. Uma estratégia para resolvê-lo e a solução esperada são discutidos.

## Sugestão

- Só pra ser super explicito você pode colocar entre parênteses que se trata de um problema de classificação binária.

"o intuito deste projeto é criar um sistema de classificação prevendo o sucesso ou não de uma solicitação (classificação binária) para imigração para os EUA."

Mas precisa mesmo? Eu, pessoalmente, fiquei com essa prática de mencionar a tarefa de aprendizagem de forma explícita pra que mesmo que o outro colega ou chefe lendo meu relatório não possa nem sonhar em interpretar errado como modelar o problema. Pro problema atual é um overkill adicionar, eu admito, mas futuramente leve isso em consideração: deixe explícito pro leitor não viajar.

As métricas utilizadas para medir o desempenho do modelo ou seus resultados são claramente definidas. A escolha de métricas é justificada com base na caracterização do projeto.

## Análise do problema

Se um conjunto de dados está disponível, atributos e estatísticas relevantes para o problema são calculados, relatados e discutidos, acompanhando uma amostra dos dados. Na ausência de um conjunto de dados, uma descrição detalhada do domínio matemático das entradas é feita. Anormalidades ou características dos dados que precisam ser tratadas são identificadas.

## Sugestão

- Se possível adicione uma amostra dos dados aqui (um snapshot de algumas features e da classe) apenas para dar uma visão de como os dados estão. Claro que isso é difícil por você ter muitas colunas, mas enfim, dar uma espiada nos dados pode ajudar a ressaltar algum ponto (por exemplo, se você mostrar uns exemplos com vários valores faltando já deixa o leitor ligado nesse aspecto). Claro que quando se trata de um problema não estruturado ou com imagens é mais óbvio o valor de incluir uma amostra :)

Uma visualização que sumariza ou extrai uma característica ou atributo relevante do conjunto de dados é criada e discutida. Indicadores visuais desse atributo são definidos com clareza.

## Excelente

- Visualizar a quantidade de nulos foi uma boa ideia!

Os algoritmos e técnicas utilizados no projeto são discutidos em detalhes e devidamente justificados com base nas características do problema.

## Sugestão

- Vou sugerir que você (em um projeto futuro ou nesse mesmo) coloque uma breve descrição do funcionamento geral dos algoritmos utilizados. Por exemplo, que o random forest é um método de ensemble que gera árvores de decisão manipulando os atributos que podem ser usados a cada split aleatoriamente etc.

É utilizado um modelo de referência claro com os quais se comparam os resultados e o desempenho das soluções desenvolvidas.

## Metodologia

Todos os passos de pré-processamento de dados são documentados com clareza. Anormalidades ou características dos dados são ajustadas conforme necessário. Se não for necessário nenhum pré-processamento, isso é justificado detalhadamente.

O processo de implementação das métricas, algoritmos e técnicas utilizados com o conjunto de dados é detalhadamente documentado. Dificuldades ocorridas durante a implementação são documentadas e discutidas.

## Excelente

- Achei interessante a forma como você organizou a explicação da aplicação (exemplo quais hiperparâmetros) e dos resultados obtidos por algoritmo.
- É aqui que se vê que você sabe o que está fazendo:

"O resultado de classificação foi extremamente ruim, com todas as classificações sendo feitas como positivas para o visto, levando ao descarte desta opção."

Ao correlacionar o resultado ruim que pode ser observado devido a utilização de outras métricas além da accuracy você demonstra domínio do problema, da solução e da análise. Parabéns :)

O processo de aprimoramento dos algoritmos e técnicas utilizados é claramente documentado. Os resultados iniciais e finais são relatados, além de passos intermediários quando cabível.

## Resultados

O modelo final, incluído seus parâmetros, é avaliado em detalhes. É feita uma análise para validar a robustez da solução encontrada.

Os resultados finais são comparados com o modelo de referência escolhido através de algum método estatístico. A conclusão de que o modelo desenvolvido é suficiente para resolver adequadamente o problema é justificada.

## Excelente

- Essa análise aqui está muito boa, além disso, fica evidente que você não apenas ficou rodando aleatoriamente, gerando resultados e plotando, mas sim você foi além e investigou a relação entre profundidade e número de membros do ensemble ( $n_{\text{estimators}}$ ).

"Inicialmente o caminho encontrado foi aumentar somente o valor de " $n_{\text{estimators}}$ ", porém, com grandes valores destes a profundidade máxima da rede tendia a crescer. Foi percebido que a melhor relação ao poder computacional vs resultado foi encontrada neste momento com valores altos para " $n_{\text{estimators}}$ ". Ao aumentar a profundidade máxima, a necessidade de poder computacional para treinamento subiu exponencialmente, porém com muito menos classificadores o resultado já era bem expressivo, sendo assim a estratégia foi modificada e passou-se a modificar a profundidade, com o learning rate ideal já encontrado, e após encontrar a melhor profundidade começou a aumentar os classificadores até encontrar o ponto ideal."

## Conclusão

Foi criada uma visualização enfatizando uma qualidade importante do projeto, acompanhada de discussão detalhada. Indicativos visuais são claramente definidos.

## Excelente

- Bom trabalho incluindo a visualização

A solução do problema de ponta a ponta é resumidamente discutida, destacando aspectos particularmente interessantes ou desafiadores do projeto.

O relatório discute como pelo menos um aspecto da implementação que possa ser aperfeiçoado. O resultado potencial desse aperfeiçoamento é comparado ou contrastado com a solução atual.

## Sugestão

- Considere também fazer um stacking das soluções obtidas, tratasse de uma técnica muito utilizada na prática para tirar proveito da saída de vários modelos. [Veja aqui](#) para mais detalhes.

## Qualidade

O relatório do projeto segue uma estrutura bem organizada e é compreensível para sua audiência. Cada seção é escrita de maneira clara, concisa e específica. O texto não contém erros gramaticais ou ortográficos em excesso. Todas as referências utilizadas na confecção do projeto são citadas.

O código é bem formatado e contém comentários para explicar implementações completas. O output do código reflete os resultados discutidos no relatório.

 [BAIXAR PROJETO](#)

RETORNAR