**Nanodegree Engenheiro de Machine Learning**

**Projeto final**

Joe Udacity  
31 de dezembro de 2050

**I. Definição**

*(aprox. 1-2 páginas)*

**Visão geral do projeto**

Nesta seção, procure fornecer uma visão de alto nível do projeto em linguagem simples. Questões para se perguntar ao escrever esta seção:

* *Foi fornecido um resumo do projeto final, bem como o domínio do problema, a origem do projeto e o conjunto de dados ou entradas?*
* *Foram dadas informações suficientes sobre o contexto para que um leitor desinformado possa entender o domínio e a enunciação problema?*

**Descrição do problema**

Nesta seção, você irá definir o problema que você está tentando resolver de forma clara, incluindo a estratégia (resumo das tarefas) que você irá utilizar para alcançar a solução desejada. Você deverá também discutir detalhadamente qual será a solução pretendida para este problema. Questões para se perguntar ao escrever esta seção:

* *A enunciação do problema foi claramente definida? O leitor irá entender o que você está esperando resolver?*
* *Você discutiu detalhadamente como irá tentar resolver o problema?*
* *A solução antecipada está claramente definida? O leitor entenderá quais resultados você está procurando?*

**Métricas**

Nesta seção, você precisará definir claramente as métricas ou cálculos que você irá usar para avaliar o desempenho de um modelo ou resultado no seu projeto. Esses cálculos e métricas devem ser justificadas baseado nas características do problema e domínio do problema. Questões para se perguntar ao escrever esta seção:

* *As métricas que você escolheu para medir o desempenho de seus modelos foram discutidas e definidas de forma clara?*
* *Você forneceu justificativas razoáveis para as métricas escolhidas, baseando-se no problema e solução?*

**II. Análise**

*(aprox. 2-4 páginas)*

**Exploração dos dados**

Nesta seção, é esperado que você analise os dados que você está usando para o problema. Esses dados podem ser tanto na forma de um conjunto de dados (ou conjuntos de dados), dados de entrada (ou arquivos de entrada), ou até um ambiente. O tipo de dados deve ser descrito detalhadamente e, se possível, ter estatísticas e informações básicas apresentadas (tais como discussão dos atributos de entrada ou definição de características das entradas ou do ambiente) Qualquer anormalidade ou qualidade interessante dos dados que possam precisar ser devidamente tratadas devem ser identificadas (tais como características que precisem ser transformadas ou a possibilidade de valores atípicos) Questões para se perguntar ao escrever esta seção:

* *Se exite um conjunto de dados para o problema em questão, você discutiu totalmente as características desse conjunto? Uma amostra dos dados foi oferecida ao leitor?*
* *Se existe um conjunto de dados para o problema, as estatísticas sobre eles foram calculadas e reportadas? Foram discutidos quaisquer resultados relevantes desses cálculos?*
* *Se****não****existe um conjunto de dados para o problema, foi realizada uma discussão sobre o espaço de entrada ou os dados de entrada do problema?*
* *Existem anormalidades ou características acerca do espaço de entrada ou conjunto de dados que necessitem ser direcionados? (variáveis categóricas, valores faltando, valores atípicos, etc.)*

**Visualização exploratória**

Nesta seção, você precisará fornecer alguma forma de visualização que sintetize ou evidencie uma característica ou atributo relevante sobre os dados. A visualização deve sustentar adequadamente os dados utilizados. Discuta por que essa visualização foi escolhida e por que é relevante. Questões para se perguntar ao escrever esta seção:

* *Você visualizou uma característica ou um atributo relevante acerca do conjunto de dados ou dados de entrada?*
* *A visualização foi completamente analisada e discutida?*
* *Se um gráfico foi fornecido, os eixos, títulos e dados foram claramente definidos?*

**Algoritmos e técnicas**

Nesta seção, você deverá discutir os algoritmos e técnicas que você pretende utilizar para solucionar o problema. Você deverá justificar o uso de cada algoritmo ou técnica baseado nas características do problema e domínio do problema. Questões para se perguntar ao escrever esta seção:

* *Os algoritmos que serão utilizados, incluindo quaisquer variáveis/parâmetros padrão do projeto, foram claramente definidos?*
* *As técnicas a serem usadas foram adequadamente discutidas e justificadas?*
* *Ficou claro como os dados de entrada ou conjuntos de dados serão controlados pelos algoritmos e técnicas escolhidas?*

**Benchmark**

Nesta seção, você deverá definir claramente um resultado de referência (benchmark) ou limiar para comparar entre desempenhos obtidos pela sua solução. O raciocínio por trás da referência (no caso onde não é estabelecido um resultado) deve ser discutido. Questões para se perguntar ao escrever esta seção:

* *Algum resultado ou valor que funcione como referência para a medida de desempenho foi fornecido?*
* *Ficou claro como esse resultado ou valor foi obtido (seja por dados ou por hipóteses)?*

**III. Metodologia**

*(aprox. 3-5 páginas)*

**Pré-processamento de dados**

Nesta seção, você deve documentar claramente todos os passos de pré-processamento que você pretende fazer, caso algum seja necessário. A partir da seção anterior, quaisquer anormalidades ou características que você identificou no conjunto de dados deverão ser adequadamente direcionadas e tratadas aqui. Questões para se perguntar ao escrever esta seção:

* *Se os algoritmos escolhidos requerem passos de pré-processamento, como seleção ou transformações de atributos, tais passos foram adequadamente documentados?*
* *Baseado na seção de****Exploração de dados****, se existiram anormalidade ou características que precisem ser tratadas, elas foram adequadamente corrigidas?*
* *Se não é necessário um pré-processamento, foi bem definido o porquê?*

**Implementação**

Nesta seção, o processo de escolha de quais métricas, algoritmos e técnicas deveriam ser implementados para os dados apresentados deve estar claramente documentado. Deve estar bastante claro como a implementação foi feita, e uma discussão deve ser elaborada a respeito de quaisquer complicações ocorridas durante o processo. Questões para se perguntar ao escrever esta seção:

* *Ficou claro como os algoritmos e técnicas foram implementados com os conjuntos de dados e os dados de entrada apresentados?*
* *Houve complicações com as métricas ou técnicas originais que acabaram exigindo mudanças antes de chegar à solução?*
* *Houve qualquer parte do processo de codificação (escrita de funções complicadas, por exemplo) que deveriam ser documentadas?*

**Refinamento**

Nesta seção, você deverá discutir o processo de aperfeiçoamento dos algoritmos e técnicas usados em sua implementação. Por exemplo, ajuste de parâmetros para que certos modelos obtenham melhores soluções está dentro da categoria de refinamento. Suas soluções inicial e final devem ser registradas, bem como quaisquer outros resultados intermediários significativos, conforme o necessário. Questões para se perguntar ao escrever esta seção:

* *Uma solução inicial foi encontrada e claramente reportada?*
* *O processo de melhoria foi documentado de foma clara, bem como as técnicas utilizadas?*
* *As soluções intermediárias e finais foram reportadas claramente, conforme o processo foi sendo melhorado?*

**IV. Resultados**

*(aprox. 2-3 páginas)*

**Modelo de avaliação e validação**

Nesta seção, o modelo final e quaisquer qualidades que o sustentem devem ser avaliadas em detalhe. Deve ficar claro como o modelo final foi obtido e por que tal modelo foi escolhido. Além disso, algum tipo de análise deve ser realizada para validar a robustez do modelo e sua solução, como, por exemplo, manipular os dados de entrada ou o ambiente para ver como a solução do modelo é afetada (técnica chamada de análise sensitiva). Questões para se perguntar ao escrever esta seção:

* *O modelo final é razoável e alinhado com as expectativas de solução? Os parâmetros finais do modelo são apropriados?*
* *O modelo final foi testado com várias entradas para avaliar se o modelo generaliza bem com dados não vistos?* -*O modelo é robusto o suficiente para o problema? Pequenas perturbações (mudanças) nos dados de treinamento ou no espaço de entrada afetam os resultados de forma considerável?*
* *Os resultados obtidos do modelo são confiáveis?*

**Justificativa**

Nesta seção, a solução final do seu modelo e os resultados dela obtidos devem ser comparados aos valores de referência (benchmark) que você estabeleceu anteriormente no projeto, usando algum tipo de análise estatística. Você deverá também justificar se esses resultados e a solução são significativas o suficiente para ter resolvido o problema apresentado no projeto. Questões para se perguntar ao escrever esta seção:

* *Os resultados finais encontrados são mais fortes do que a referência reportada anteriormente?*
* *Você analisou e discutiu totalmente a solução final?*
* *A solução final é significativa o suficiente para ter resolvido o problema?*

**V. Conclusão**

*(aprox. 1-2 páginas)*

**Foma livre de visualização**

Nesta seção, você deverá fornecer alguma forma de visualização que enfatize uma qualidade importante do projeto. A visualização é de forma livre, mas deve sustentar de forma razoável um resultado ou característica relevante sobre o problema que você quer discutir. Questões para se perguntar ao escrever esta seção:

* *Você visualizou uma qualidade importante ou relevante acerca do problema, conjunto de dados, dados de entrada, ou resultados?*
* *A visualização foi completamente analisada e discutida?*
* *Se um gráfico foi fornecido, os eixos, títulos e dados foram claramente definidos?*

**Reflexão**

Nesta seção, você deverá resumir os procedimentos desde o problema até a solução e discutir um ou dois aspectos do projeto que você achou particularmente interessante ou difícil. É esperado que você reflita sobre o projeto como um todo de forma a mostrar que você possui um entendimento sólido de todo o processo empregado em seu trabalho. Questões para se perguntar ao escrever esta seção:

* *Você resumiu inteiramente o processo que você utilizou neste projeto?*
* *Houve algum aspecto interessante do projeto?*
* *Houve algum aspecto difícil do projeto?*
* *O modelo e solução final alinham-se com suas expectativas para o problema, e devem ser usadas de forma geral para resolver esses tipos de problemas?*

**Melhorias**

Nesta seção, você deverá discutir como um aspecto da sua implementação poderia ser melhorado. Por exemplo, considere maneiras de tornar a sua implementação mais geral e o que precisaria ser modificado. Você não precisa fazer a melhoria, mas as possíveis soluções que resultariam de tais mudanças devem ser consideradas e comparadas/contrastadas com a sua solução atual. Questões para se perguntar ao escrever esta seção:

* *Existem melhorias futuras que podem ser feitas nos algoritmos ou técnicas que você usou neste projeto?*
* *Existem algoritmos ou técnicas que você pesquisou, porém não soube como implementá-las, mas consideraria usar se você soubesse como?*
* *Se você usou sua solução final como nova referência, você acredita existir uma solução ainda melhor?*