

Visão geral do projeto de conclusão

Neste projeto final, você vai se apoiar em todo o aprendizado obtido durante o programa Nanodegree para solucionar um problema de sua escolha ao aplicar algoritmos e técnicas de machine learning. Primeiro, você **definirá** o problema que quer resolver e investigará potenciais soluções e execução de métricas. Depois, **analisará** o problema através de visualizações e exploração de dados para ter um melhor entendimento de quais algoritmos e características são apropriados para resolvê-lo. Então, **implementará** o algoritmo e métricas escolhidos, documentando os passos de pré-processo, refinamento e pós-processo ao longo do caminho. Depois, você coletará **resultados** sobre o desempenho dos modelos usados, visualizará quantidades significantes e validará/justificará esses valores. Por último, você descreverá suas **conclusões** sobre os resultados e discutirá se sua implementação soluciona o problema adequadamente.

Destaques do projeto

Este projeto é desenvolvido para preparar você para entregar um relatório de solução elegante e completo sobre um problema do mundo real em uma área de interesse. Quando desenvolvemos novas tecnologias ou derivamos adaptações de tecnologias anteriores, documentar apropriadamente seu processo é fundamental tanto para validação como para replicar seus resultados.

O que você aprenderá ao concluir este projeto:

- Como pesquisar e investigar um problema de interesse do mundo real.
- Como aplicar precisamente algoritmos e técnicas de machine learning específicos.
- Como analisar e visualizar adequadamente seus dados e resultados para validação.
- Como documentar e escrever um relatório sobre seu trabalho.

Descrição do projeto

Pense sobre o domínio de um problema pelo qual você tenha paixão, como assistência médica, engenharia, finanças, robótica, marketing, bioinformática ou qualquer outra área que motive você. Escolha um problema existente dentro desse domínio e que você possa resolver ao aplicar algoritmos e técnicas de machine learning. Certifique-se de coletar todos os recursos necessários (como conjunto de dados) para concluir este projeto e fazer as citações apropriados quando necessário.

Para qualquer área de aplicação ou problema que você está investigando, existem cinco grandes estágios que você irá percorrer para este projeto final e, posteriormente, documentar. Cada estágio atua em um papel significativo dentro do ciclo da vida de desenvolvimento, iniciando com a definição de um problema e finalizando com uma solução de trabalho elegante. Como você irá construir seu próprio caminho desenvolvendo seu projeto, tenha certeza de que você está trabalhando também em um rascunho de seu relatório, já que esse é o aspecto mais importante do projeto!

Para determinar se seu projeto, domínio do problema e o problema que você quer resolver se encaixam na visão da Udacity sobre o projeto final de machine learning, veja a [rubrica do projeto](#) e anote cada critério da rubrica no qual você será avaliado. Um projeto satisfatório terá um relatório que incorpora cada estágio e componente da rubrica.

Requisitos de software

Seu projeto deve ser escrito em Python 2.7 ou 3.x. Dada a forma livre do projeto de conclusão de machine learning, o software e as bibliotecas que você usará para concluir seu projeto com êxito vão variar dependendo da área de aplicação e do problema escolhido. Por causa disso, é mandatório que todo software e biblioteca necessários utilizados em seu projeto final sejam acessíveis para o revisor e claramente documentados. Informações sobre software e bibliotecas que seu projeto utiliza devem ser incluídas no README, junto à sua submissão. Note que software proprietário, software que precisa de licença privada ou software que está atrás de uma paywall ou login devem ser evitados.

Requisitos de dados

Todo o projeto de conclusão de machine learning certamente precisará de alguma forma de conjunto de dados ou entrada de estrutura de dados (entrada de arquivos de texto, imagens, etc.). Assim como os requisitos de software acima, os dados que você utilizará devem ser publicamente acessíveis ou fornecidos por você durante o processo de envio, e dados proprietário ou privados não devem ser utilizados sem permissão expressa. Leve em consideração o tamanho do arquivo dos seus dados - enquanto não há restrição de limite máximo, entradas de arquivos que são maiores que 200MB podem levar mais de um minuto para serem obtidos pelos revisores. Isso pode fazer com que leve mais tempo para que o projeto seja avaliado. Se os dados em que você está trabalhando são muito pesados para serem transferidos, considere se você pode trabalhar com apenas um conjunto de dados.

Ética

O teste A/B do curso da Udacity, como parte do Nanodegree Analista de Dados, tem um segmento que discute [a sensibilidade dos dados](#) e a expectativa de privacidade daqueles cuja informação foi coletada. Enquanto a maioria dos dados que você encontra disponíveis para o público geral não têm complicações éticas, é extremamente importante que você considere de onde os dados que você está utilizando vêm e se esses dados não contêm nenhuma informação delicada. Por exemplo, se você trabalha para um banco e quer usar declarações de seus usuários como parte do seu projeto, essa provavelmente seria uma escolha não ética de dados, e deve ser evitada.

Se você tem alguma dúvida em relação à natureza dos dados ou software que você pretende usar para o projeto final, envie um e-mail para ml-suporte@udacity.com, com o assunto "*Pesquisa de dados/software para projeto final*".

Orientações para o relatório

Seu projeto será avaliado principalmente pelo relatório que será enviado. Espera-se que o relatório do projeto contenha detalhes suficientes, documentação, análise e uma discussão que reflita adequadamente o trabalho que você realizou neste projeto. Por causa disso, é extremamente importante que o relatório seja escrito de maneira profissional e padronizada, de forma que aqueles que revisam seu projeto sejam capazes de identificar claramente cada componente do projeto no relatório. Sem um relatório escrito adequadamente, seu projeto não pode ser avaliado da melhor maneira. Um [modelo de relatório de projeto](#) é fornecido para que você entenda como um relatório de projeto deve ser estruturado. Nós incentivamos nossos alunos a fazerem um **relatório de aproximadamente nove a quinze páginas**.

O relatório do projeto de final de machine learning não deve ser tratado diferente de um artigo de pesquisa acadêmica. Seu objetivo é, em última análise, apresentar a pesquisa que você desenvolveu dentro do domínio do respectivo problema que escolheu e, então, discutir cada estágio do projeto até ele ser completado. A narrativa encontrada no [modelo de relatório de projeto](#) fornece um "*checklist do projeto*", que auxiliará você a se manter no caminho certo para ambos o projeto e a documentação em seu relatório. Cada estágio pode ser encontrado como uma seção que guiará você através de cada componente do ciclo de vida de desenvolvimento do projeto. Faça bom uso desses recursos!