

# Plano de Avaliação - Dell Academy

## Trilha de Machine Learning

### Descrição da atividade

Desenvolver um projeto de Machine Learning para análise de sentimento de opiniões sobre filmes. O dataset escolhido para a execução desta atividade é o [IMDB 50K Movie Reviews](#). As mentoradas deverão entregar a atividade em um notebook ou conjunto de scripts executáveis. O projeto deve conter todas as etapas estudadas durante o período de mentoria: análises e inferências exploratórias dos dados, pré-processamento, treinamento e avaliação do classificador desenvolvido, e visualização dos resultados obtidos. As mentoradas serão encorajadas a colocar em prática todas as técnicas aprendidas e usar seu senso crítico e analítico para decidir qual a abordagem irá oferecer melhores resultados.

### Avaliação

Diversos aspectos observados e avaliados neste projeto. Tais aspectos englobarão tanto a capacidade de apresentar e interpretar dados, como também o pré-processamento, treinamento e avaliação do modelo utilizado na solução.

#### Análise Exploratória (10 pontos)

O projeto deverá conter, no mínimo, uma análise básica das métricas estatísticas dos dados, tais como: distribuição dos dados, máximos e mínimos, médias, medianas e desvios padrões. Será esperado também no mínimo um gráfico simples de estatística descritiva, podendo ser: barras, histograma, dispersão, entre outros. Espera-se também que alguma inferência a respeito dos dados seja obtida a partir dessas métricas.

#### Desempenho do Modelo (50 pontos)

Será levada em consideração a capacidade de generalização do modelo, que deverá ser capaz de classificar corretamente o sentimento presente nas avaliações de um filme. Serão observadas, para fins avaliatórios, as seguintes características do modelo:

- Pré-processamento dos dados;
- Escolha e implementação do algoritmo;
- Escolha de valores utilizados como hiperparâmetros para criação do modelo;
- Acurácia do modelo sobre o conjunto de testes fornecido no dataset;

- Escolha e interpretação de métricas de avaliação do modelo;

## Visualização das informações (30 pontos)

Os resultados obtidos, bem como o modelo, devem ser sumarizados e apresentados de forma legível e de fácil compreensão. Os quesitos que serão avaliados para compor a pontuação total desta atividade serão:

- Escolha dos gráficos (barra, histograma, linhas, etc.);
- Escolha de cores adequadas;
- Gráficos Representativos.

## Apresentação (10 pontos)

Uma apresentação do projeto deverá ser realizada, onde será explicado o passo a passo da solução desenvolvida. Além disso, serão separados alguns textos além dos fornecidos no dataset para serem classificados durante a apresentação. Dessa forma, será possível avaliar o domínio sobre o que foi desenvolvido pelas mentoradas.