Relatório Projeto 3 de Ciência de Dados

**Rating de filmes (IMDB)**

por: Marcelo Lisboa, Tiago Bergamo e Bruno Kaczelnik - 2C Engenharia Insper

**Divisão do Trabalho:**

Encontramos dois datasets, um muito maior, muito mais confuso, que precisaria de muita atenção para então ser utilizado no projeto e outro menor, mais fácil de utilizar, porém com menos dados. Assim, resolvemos deixar um dos integrantes para cuidar do dataset mais complexo enquanto os outros dois começariam já o projeto.

O integrante Bruno foi o responsável em melhorar esse dataset o máximo possível para torná-lo usável. Contudo, pelo curto período de tempo, infelizmente não conseguiu terminar. Além disso, Bruno também procurou quais seriam os modelos que usaríamos para classificar os dados e contribuiu em algumas partes das análises e da construção do código do outro dataset.

O integrante Marcelo foi o responsável pela análise exploratória do modelo. Junto com o integrante Tiago, ambos se aprofundaram no modelo, fazendo a seleção de variáveis, primeiro com os atores e diretores, depois com os gêneros, com os dois juntos e por fim com todos os indicadores escolhidos. Com isso pronto, Tiago aplicou os dados nos classificadores selecionados.

Os três integrantes trabalharam na obtenção da resposta de um novo filme no dataset. Marcelo e Tiago ainda fizeram uma matriz de confusão com o intuito de explicar a baixa acurácia nas respostas do modelo Multinomial Naive Bayes, inclusive sobre o resultado do filme testado (Avengers: Infinity War).

Para concluir, juntos, todos fizeram o relatório final do projeto.