

**Projeto 3 Ciência dos Dados - Diário de Progresso**

*Guilherme Lunetta*

*Arthur Carvalho*

*Fernando Bichuette*

*Rafael Kahn*

Sala 2C – Engenharia

São Paulo, 31 de outubro de 2019

**Dia 31/10 – Kickoff do Projeto**

Primeiro dia para falar do projeto, começamos pensando em temas de interesse do grupo para iniciar as pesquisas sobre qual método usar. Chegamos em dois temas, Esportes Eletrônicos e Mercado Imobiliário. Fora do Insper, procuramos por Database’s e encontramos um sobre Imóveis em São Paulo, de abril de 2019, bem recente e com mais de 10000 imóveis, ou seja, um database muito bom de se trabalhar.

**Dia 5/11 – Definição de grupo e 2 propostas de tema por grupo**

Nesse dia, passamos para a professora Kelly os grupos e começamos a trabalhar no documento a ser entregue no final do dia. O documento consiste em 2 propostas de projeto, no nosso caso, uma dizia que trabalharíamos com regressão para predizer o valor de um imóvel. Já na segunda proposta, trabalharíamos com classificador para classificar imóveis em relação a luxuosidade, com 3 tipos, Alto, Médio e Baixo.

**Dia 7/11 e 12/11**

No dia 7, foi iniciada a análise exploratória da nossa base de dados, dando enfoque na variável do preço. Assim, plotamos gráficos do preço por qualquer outra variável relevante para o desenvolvimento do nosso projeto. Chegando no dia 12/11, concluímos tal análise exploratória, assim como adicionamos outras informações relevantes ao projeto, como markdowns explicando o nosso raciocínio e filtros de informações para apenas analisarmos o que for relevante para nós. Além disso, decidimos ao fim dessa aula que iremos utilizar o método do classificador random forest para o futuro do trabalho.

**Dia 14/11**

Nesse dia, além de termos iterados sobre diversos aspectos do projeto, também fizemos completamente o classificador random forest, com uma acurácia total de 70%. Decidimos passos futuros para o projeto e, por fim, começamos a desenvolver uma matriz de confusão para os resultados do nosso classificador.

**Dia 19/11**

Na aula do dia 19, foi feita a matriz de confusão por completa e plotados outros gráficos. Além disso, organizamos o nosso repositório com informações faltantes e limpamos arquivos inúteis. Começamos a ver outros métodos de análise que podem ser úteis para nosso trabalho.