

# Relatório de Conclusão

## Aprendizado de Máquina

**Professor:** Heraldo Almeida

**Aluno:** Leonardo Freire

**Kaggle Id:** lfreire80

### Primeiro passo

Meu primeiro paço em busca do objetivo foi tentar implementar os algoritmos mais simples visto em sala de aula, iniciando com uma regressão linear simples, e posteriormente uma regressão polinomial utilizando todos os atributos.

Os resultados foram desastrosos com  $R^2$  muito inferiores a zero.

Após análise dos dados para treinamento identifiquei algumas discrepâncias nos maiores e menores valores dos imóveis, com isso resolvi eliminá-los.

Após a limpeza o resultado de  $R^2$  passou para valores maiores que zero.

## **Segundo passo**

Na tentativa de encontrar os melhores atributos para o modelo, fiz alguns gráficos onde cruzava o valor do atributo com o valor do imóvel.

Com isso identifiquei alguns potenciais atributos.

- quartos
- suites
- vagas
- area\_util,
- piscina,
- ginastica,
- vista\_mar

## **Terceiro passo**

Apos algumas tentativas ajustando o grau do polinômio ainda não estava obtendo resultados satisfatórios com a regressão linear. Com isso implementei outros algoritmos de Regressão na tentativa de melhores resultados.

- Floresta aleatória
- KNN (K-vizinhos mais próximos)
- (SVM) Support Vector Machine
- Árvore de decisão
- Gaussian Naive Bayes
- K-Means

## Conclusão

Dentre os algoritmos testados, o que obteve melhores resultados foi a Floresta aleatória, que após os ajustes, possibilitou um  $R^2$  por volta de 0.70 na amostra de teste. Que refletido no resultado final num score de 0.30513 no Kaggle

Os parâmetros utilizados no Random Forest foram

- max\_depth=7
- random\_state=6
- n\_estimators=40

Todos os testes foram implementados no script

`./src/descoberta.py`

A execução da massa de teste e geração do resultado final encontra-se no script:

`./src/execucao_RandomForest.py`