

**Desafio - Parte 2**  
**Mineração de dados - S73**  
**Aluno: Luan Carlos Klein**

**Breve descrição das estratégias utilizadas**

A estratégia utilizada para a predição se um estabelecimento é destaque ou não pode ser separada em três partes distintas.

**Parte 1 - Definição das features**

Nessa parte, foram utilizadas algumas estratégias para a definição das features, e com isso, 4 categorias de features foram adicionadas.

*Feature 1* - Sequências frequentes de categorias que implicam em ser ou não destaque.

- Utilizou-se algoritmos para encontrar as categorias e/ou conjuntos de categorias que tendem a implicar em um estabelecimento ser ou não destaque.

*Feature 2* - Sequências frequentes de atributos que implicam em ser ou não destaque.

- Da mesma forma que na feature 1, porém, agora voltado para os atributos dos estabelecimentos.

*Feature 3* - Sentimento médio das avaliações.

- Utilizou-se um algoritmo para avaliar o sentimento por detrás de cada um dos reviews, e foi feito a média para cada estabelecimento. Estabelecimentos sem avaliações, foram considerados neutros.

*Feature 4* - Proximidade de lugares de destaque.

- Utilizou-se o algoritmo DBSCAN para encontrar lugares com alta densidade de estabelecimentos destaques e não destaques, e cada um dos estabelecimentos foi classificado se fazia parte de um desses lugares de alta densidade. Após, foi feita a 'predição' dos lugares de teste, ou seja, avaliado se um estabelecimento fazia parte daquelas zonas densas ou não.

Por fim, também foram utilizados os dados de localização (latitude e longitude), que foram transformados para ficar entre 0 e 1.

**Parte 2 - Estratégias de Predição**

Três modelos distintos foram criados e treinados com os dados criados acima. São eles:

- 1 - Floresta aleatória;
- 2 - Regressão Múltipla;
- 3 - Rede Neural.

### **Parte 3** - Democracia das estratégias de predição

Por fim, para realizar a predição, os 3 modelos foram aplicados nos dados de teste, e para melhorar o resultado, utilizou-se uma espécie de votação para cada um dos lugares preditos pelos modelos, ou seja, para um lugar ser definido como destaque, seria necessário que pelo menos dois dos três modelos previssem que tal lugar é destaque.