**Trabalho final classificação/regressão (18/11/2020)**

Os datasets contém dados históricos de vendas para 1.115 lojas. A tarefa é prever a coluna "Vendas" no conjunto de teste. Observe que algumas lojas no conjunto de dados foram temporariamente fechadas para reformas.

**Descrição dos Arquivos**

* **dataset\_treino.csv** - dados históricos, incluindo vendas
* **dataset\_teste.csv** - dados históricos excluindo vendas
* **sample\_submission** - contém o template para você enviar a sua solução para mim
* **lojas.csv** - informações suplementares sobre as lojas

## Descrição dos Campos

* **ID** - um ID que representa uma tupla (Store, Date) dentro do conjunto de dados
* **Store** - um ID único para cada loja
* **Sales** - o volume de negócios de um determinado dia (é isso que você está prevendo)
* **Customers** - o número de clientes em um determinado dia (essa informação não está presente no conjunto de teste)
* **Open** - um indicador para saber se a loja estava aberta: 0 = fechada, 1 = aberta
* **StateHoliday** - indica um feriado estadual. Normalmente, todas as lojas, com poucas exceções, estão fechadas nos feriados estaduais. Note que todas as escolas estão fechadas nos feriados e fins de semana. a = feriado público, b = feriado de Páscoa, c = Natal, 0 = Nenhum
* **SchoolHoliday** - indica se a (Store, Date) foi afetada pelo fechamento de escolas públicas
* **StoreType** - diferencia entre 4 modelos de lojas diferentes: a, b, c, d
* **Assortment** - descreve um nível: a = básico, b = extra, c = estendido
* **CompetitionDistance** - distância em metros até a loja concorrente mais próxima
* **CompetitionOpenSince[Month/Year]** - indica o ano e mês aproximado da hora em que o concorrente mais próximo foi aberto
* **Promo** - indica se uma loja está executando uma promoção nesse dia
* **Promo2** - promoção contínua e consecutiva para algumas lojas: 0 = loja não está participando, 1 = loja está participando
* **Promo2Since[Year/Week]** - descreve o ano e a semana do calendário em que a loja começou a participar do Promo2
* **PromoInterval** - descreve os intervalos consecutivos que o Promo2 é iniciado, indicando os meses em que a promoção é iniciada novamente. Por exemplo. "Feb, May, Aug, Nov" significa que cada rodada começa em fevereiro, maio, agosto, novembro de qualquer ano para aquela loja.

Vocês podem utilizar tudo que aprenderam na disciplina até o momento (normalização, seleção e construção de atributos, balanceamento, dentre outros) e também procurar outras técnicas que não foram vistas. **Na aula de quarta (25/11/2020) cada um deve apresentar o que usou para resolver o problema (em até 10 minutos), mas antes, quero que me enviem um csv no formato do arquivo sample\_submission contendo as predições do conjunto de teste. Irei utilizar a métrica “Score” para definir qual o método que produziu o melhor resultado. Após todas as apresentações irei divulgar qual o método que produziu o melhor resultado.**

**Importante: A nota será atribuída em função de tudo que o aluno realizou, e não apenas do valor de Score obtido.**

**Importante2: Vocês podem conversar entre si e se ajudar, mas lembrem-se que estão competindo (HAHAHA).**