

ATIVIDADE INTEGRADORA

Como cientista de dados de um grande conglomerado, você recebeu as seguintes demandas de diferentes áreas da empresa:

- criar um sistema de recomendação que indique algumas opções de estabelecimentos com base no histórico de avaliações positivas dos usuários. Por exemplo, sabe-se que usuários que avaliam positivamente o estabelecimento A, geralmente, avaliam positivamente o estabelecimento B. Assim, para um cliente que já avaliou positivamente o estabelecimento A, pode-se recomendar o estabelecimento B.
- suspeita-se que a entrega de cupons de desconto possui impacto positivo nas avaliações recebidas por um estabelecimento. Boas avaliações incluem não apenas boas notas, mas também alto impacto, caracterizado pelo número de classificações da avaliação como útil (useful) e pela rede de relacionamento do revisor. Portanto, uma estratégia inteligente de CRM envolve direcionar cupons para aqueles clientes com maior probabilidade de fazer uma boa avaliação. Seu trabalho é escolher um estabelecimento de sua preferência e, analisando o histórico de avaliações, construir um algoritmo capaz de recomendar quais usuários do Yelp da região do estabelecimento devem receber cupons, com o objetivo de maximizar as boas avaliações para o estabelecimento. Caso seja necessário, pode-se considerar que pessoas que citaram o termo cupom em suas avaliações receberam um cupom do estabelecimento sendo revisado.

Após alguma negociação, uma dessas entregas ficará prevista para o seu próximo sprint considerando a entrega de um relatório. Esse relatório deverá conter a descrição dos passos técnicos, justificativa das principais escolhas, análises e conclusões. Adicionalmente, a partir dos resultados obtidos, apresente uma recomendação para o board de diretores (as) - procure condensar a ideia, bem como os argumentos, de forma simples e objetiva, tal indicação deve ser feita em no máximo 5 linhas.



Para executar a atividade, você deverá fixar uma região para trabalhar a partir da base de dados Yelp (<https://www.yelp.com/dataset>).

Entrega

- Apresentação em aula dos resultados;
- **Computação em Larga Escala:** jupyter notebook executável e sem erros no computador em nuvem da disciplina, comentado com o passo a passo do algoritmo desenvolvido incluindo todas as etapas, desde a análise até a preparação dos dados para a etapa de modelagem;
- **Modelagem Preditiva Avançada:** relatório explicando a análise e justificando as escolhas utilizadas (o código R deverá ser entregue a parte ou, se preferir, incorporado no relatório).

Datas

Apresentação e entrega do relatório: **04/07, sábado, no período da manhã (08 – 12 horas)**.

A ordem das apresentações será definida por sorteio no próprio dia.

Cada apresentação terá 30 minutos, sendo 20 minutos de apresentação e 10 minutos de dúvidas.

A nota final combinará a utilização adequada das ferramentas e métodos apresentados em aula, qualidade da análise, criatividade, clareza na apresentação e explicação das dúvidas.