

Projeto Final de Machine Learning II

Objetivos. Com o ferramental em análise e em ciência de dados adquirido até o momento, a ideia é que, com este projeto, vocês possam avaliar um dataset sob três óptica: (1) uma análise exploratória de dados; (2) uma análise de segmentação; e (3) uma análise preditiva.

Dataset. O conjunto de dados a ser utilizado para desenvolver este projeto é de livre escolha do grupo. Entretanto, dados os requisitos (2) e (3) mencionados acima, a base de dados deve possuir uma variável que deverá ser predita pelos modelos que vocês desenvolverão.

Proposta do projeto. Para atingir os objetivos do projeto, vocês deverão entregar um **notebook**, que contemple as três etapas supramencionadas. Especificamente, espera-se que:

- na etapa de análise exploratória de dados, o grupo seja capaz de entender o problema de negócio ao qual o dataset escolhido se insere, sabendo levantar perguntas de interesse, que deverão ser respondidas com embasamentos nas análises de dados (avaliações gráficas, estatísticas e afins);
- para a avaliação de segmentação, espera-se que o grupo seja capaz de encontrar agrupamentos no conjunto de dados, em relação a variáveis de interesse para o problema proposto. Por exemplo, pensando em um dataset de compras de determinados produtos, é possível associar perfis de consumidores a padrões específicos de gastos? Como isso pode ser feito? (Apenas um exemplo; a ideia é que vocês consigam transpor esse raciocínio ao dataset de vocês!);
- por fim, a análise preditiva pressupõe a utilização de modelos, a escolha do grupo, para prever a(s) variável(eis) de interesse dos problemas levantados pelo grupo. Apesar de ser interessante vocês praticarem os algoritmos estudados neste módulo, comparações com modelos previamente estudados são sempre bem-vindas!

Entregável. Notebook do projeto (arquivo ou github).

Preenchem a [planilha de divisão de grupos do projeto final](#).

Quaisquer dúvidas, estou sempre à disposição! :)