

## Mini-Projeto 2 - Introdução à Programação

### Introdução à Ciência de Dados

Base de dados sugerida para testes durante a codificação:

<https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data>

Base de dados que será usada na correção: *outras*

- (a) Crie a classe *ExtratorDeProbabilidades* que contém ao menos os seguintes atributos (i) referência a um arquivo csv, e (ii) uma estrutura de lista com tuplas e dicionários para armazenar os dados da base.
- (b) Crie um método *carregar\_colunas(lista\_colunas, quantidade)* que recebe uma lista de strings (representando o nome das colunas da base) e um número inteiro (representando uma quantidade de registros), e carrega uma amostra dos registros de tamanho *quantidade* com apenas as colunas em *lista\_colunas*. **Atenção:** os dados carregados deverão ser codificados na estrutura de lista que é atributo da classe.
- (c) Crie o método *descarregar()* que reinicializa a estrutura de lista como vazia.

- (d) Crie o método *probabilidade\_apriori((característica, valor))* que calcula - para os dados carregados em memória - a probabilidade da coluna *característica* possuir o valor *valor*.

ex.:

$$P(\text{modelo} == \text{'celta'}) = \frac{\#(\text{modelo} == \text{'celta'})}{\#\text{carros}}$$

- (e) Crie o método *probabilidade\_apriori\_intervalo((característica, (início,fim)))* que permite o cálculo - para os dados carregados em memória - da probabilidade a priori considerando intervalos numéricos:

ex.:

$$P(0 < \text{odometro} < 10000) = \frac{\#(0 < \text{odometro} < 10000)}{\#\text{carros}}$$

- (f) Crie o método *probabilidade\_condicional(((característica\_1, valor\_1),(característica\_2, valor\_2)))* que calcula - para os dados carregados em memória - a probabilidade condicional da coluna *característica\_1* possuir o valor *valor\_1* dado que *característica\_2* possui o valor *valor\_2*:

ex.:

$$P(\text{modelo} == \text{'celta'} \mid \text{cambio} = \text{'manual'}) = \frac{\#(\text{modelo} == \text{'celta'} \text{ and } \text{cambio} == \text{'manual'})}{\#(\text{cambio} == \text{'manual'})}$$

- (g) Crie o método *probabilidade\_apriori\_intervalo((característica, valor), (característica, (início, fim)))* ou *probabilidade\_apriori\_intervalo((característica, (início, fim)), (característica, valor))* ou *probabilidade\_apriori\_intervalo((característica, (início, fim)),*

(*característica, (início, fim)*) que permite o cálculo - para os dados carregados em memória - da probabilidade condicional considerando intervalos numéricos:

$$P(0 < valor < 50000 | cambio = 'manual') = \frac{\#(0 < valor < 50000 \text{ and } cambio == 'manual')}{\#(cambio == 'manual')}$$

- (h) Integre todas as funcionalidades em um Menu a partir do qual é possível carregar e descarregar bases de dados, consultar todas as probabilidades apriori e condicionais.

### DESAFIO (opcional):

Dados os valores para ***n* características observadas, uma característica alvo**, e um **nível de probabilidade**, indique quais valores para alvo atende à probabilidade mínima desejada.

Ex.:

*Qual o modelo de carro cuja a probabilidade de ser encontrado é maior que 90% dado que o seu preço é menor que R\$40.000,00 e o seu câmbio é manual?*

$$P(\text{modelo} == ? | valor < 40000 \text{ and } cambio == 'manual') > 0.9, \text{modelo} = \{'celta', 'gol'\}$$

*Qual o intervalo de valores cuja a probabilidade de se encontrar um carro é superior a 90% dado que o modelo é celta e o câmbio manual?*

$$P(? < valor < ? | modelo == 'celta' \text{ and } cambio == 'manual') > 0.9, 0 < valor < 30.000$$