

 <p>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA PIAUÍ</p>	<p><b>INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ</b></p> <p><b>Curso: ADS</b></p> <p><b>Disciplina: Tópicos Especiais em Sistemas de Informação</b></p> <p><b>Professor: Ely</b></p>
--	--

### Exercício 04

- 1) Com o arquivo carregue o arquivo risco-credito.csv.
- 2) Use `exiba` as 3 primeiras linhas do dataset.
- 3) Exiba dados do dataset usando a função `describe()`.
- 4) Exiba os valores únicos de todas as colunas.
- 5) Filtre todos os registros que possuem risco alto e depois os que possuem risco baixo.
- 6) Verifique se alguma coluna do dataset possui valores nulos.
- 7) Separe as colunas do dataset em dataframes menores `x` (dados) e `y` (marcações). A coluna `y` deve ser a 'risco', as colunas que compõem o `x` são as demais.
- 8) Crie variáveis categóricas para o dataframe `x` e exiba o resultado da conversão.

O dataframe `y` possui uma variável do tipo categórica. na verdade, chamamos de um dataset com multiclassificação. Há algoritmos apropriados para mais de 2 resultados, como nessa situação. Não utilizaremos um algoritmo de multiclassificação neste momento. Vamos aplicar o nosso algoritmo da forma como está e ver o seu desempenho.

- 9) Separe as variáveis de treino e teste com 25% dos registros para teste.
- 10) Crie um modelo para o algoritmo MultinomialNB e o execute. Qual o score obtido? Foi baixo?
- 11) Vamos agora fazer uma pequena alteração. Vamos classificar aqueles cujo risco é moderado para alto com o código abaixo:

```
y = y.map( {'alto': 1, 'baixo': 0, 'moderado': 1} )
```

Obtenha novamente os dados de treino e teste e veja o novo score. Faça a mesma alteração alterando 'moderado' para risco baixo.

A performance do algoritmo melhorou?