

Certamente seu celular está próximo, ou agora está se perguntando onde ele está. O smartphone nos permite conectividade com a Internet o dia todo, utilizar serviços de localização, comunicação em vídeo, voz e texto e trabalhar remotamente com equipes globais. Para isso precisamos do dispositivo e, é claro, da conexão 3G ou 4G.

Como parte deste serviço fundamental, as operadoras têm um papel estratégico importante e complexo para garantir esta conectividade final (*last mile*) entre nosso dispositivo e a Internet.

Nesta atividade iremos analisar dados de entrevistas de uma pesquisa de satisfação e qualidade percebida, relativa à prestação de Serviço Móvel pessoal (telefonia móvel e banda larga móvel), na modalidade pré-paga (fonte dos dados e questionário da pesquisa: <http://dados.gov.br/dataset/banco-de-dados-da-pesquisa-telefonia-movel-pre-paga>).

Considerando a seguinte questão:

**J1: Nível de satisfação geral do entrevistado com a prestadora citada, levando em conta toda a experiência com esta**

	Totalmente Insatisfeito	Satisfeito									Totalmente Satisfeito	NS/ NR
Satisfação geral com a prestadora	0	1	2	3	4	5	6	7	8	9	10	99

trabalharemos com duas classes de clientes. A primeira será a classe de clientes com satisfação **alta** (nível de satisfação maior ou igual a 8) e a segunda classe será a de clientes com satisfação **baixa** (nível de satisfação menor que 8). O **objetivo** da atividade consiste em criar um classificador de satisfação (alta ou baixa) de clientes com bom desempenho preditivo. Para isso você deverá fazer a organização, limpeza e análise gráfica no Python. A modelagem preditiva e avaliação dos modelos aplicados a

# Inspêr

esses dados deverá ser feita no R. Você pode priorizar o erro de classificação de forma geral ou evitar falsos positivos/falsos negativos, desde que indique claramente o critério no relatório. Além disso, se achar oportuno, você pode combinar os dados dessa pesquisa com outra base de dados para melhorar a performance preditiva do seu classificador.

## Entrega

- **Computação para Ciência de Dados:** deverá ser feita por um Jupyter Notebook com todos os passos utilizados no processamento dos dados e também descrição sobre cada passo em markdown
- **Modelos Preditivos:** relatório explicando a análise e justificando as escolhas dos modelos (o código R deverá ser entregue a parte ou, se preferir, incorporado no relatório)

## Datas

Será realizada entrega no Blackboard e apresentação do projeto no próprio Jupyter Notebook e R no último dia de aula.

## Extra

Em adicional podemos também combinar os dados de outra pesquisa com operadoras da Anatel que são relacionadas, conforme segue:

<https://www.anatel.gov.br/paineis/consumidor/pesquisa-de-satisfacao>

Os dados podem ser baixados do nosso Google Drive, mas os dados das fontes originais podem estar mais atualizados. Encontre formas de identifica uma chave para combinar os dados destas fontes diferentes.

<https://drive.google.com/file/d/1NIFCAh7eYS6x4nE4wo8RXOeMzby5TUgm/view?usp=sharing> (~410MB)