

# Projeto | Inteligência na Web e Big Data

## Utilizando *Naive Bayes* para estimar a probabilidade dos comentários escolhidos pelo editor

Adriana Cavalcanti

- Seleção automática dos comentários
- Verificação da imagem a ser passada  
(clássica vs ousada)
- Comparação de imagem entre as empresas
- *New York Times Comments* Abril/2018
- *Framework Spark*
- *Linguagem Python*
- Classificador *Naive Bayes*

# WorkFlow

Sample & Clear  
DATAFRAME

Selection  
FEATURES

Bag of Words  
COMMENTS

TakeOrdered(10)  
WORDS

OHE

*Naive Bayes*

# Resultados

Sample & Clear  
DATAFRAME

Selection  
FEATURES

Bag of Words  
COMMENTS

TakeOrdered(10)  
WORDS

OHE

*Naive Bayes*

- Tratamento de OHE no DataPrep
- Inserir *Stop Words*