

# RECUPERAÇÃO DA INFORMAÇÃO NA WEB E EM REDES SOCIAIS

*Curso de Especialização em Ciência de Dados (Big Data Processing and Analytics)*

Prof. Augusto Lozano



## Avaliação Final

- Os alunos deverão utilizar a empresa elaborada em outras disciplinas e apresentando um problema relacionado a processamento de língua natural

## Avaliação Final

- Conteúdo:
  - Classificação de texto
  - Fonte de dados: scrapping ou conjunto de dados disponíveis na web
  - Gráfico de nuvem de palavras com e sem aplicação de stop words
  - Gráfico mostrando o balanceamento do conjunto de dados
  - Vetor word2vec com similaridade de 3 palavras
  - Apresentar as técnicas de pré-processamento aplicadas: tokenização, stemming, lemmatization, POS tag

