

## Bootcamp: Arquiteto(a) de Machine Learning

### Trabalho Prático

#### Módulo 2: Processamento de Linguagem Natural

#### Objetivos de Ensino

Exercitar os seguintes conceitos trabalhados no Módulo:

1. Conhecer alguns métodos relacionados às principais bibliotecas de Processamento de Linguagem Natural com Python.
2. Praticar técnicas de pré-processamento textual com base em citações de autores.
3. Analisar conteúdo textual a partir da aplicação de técnicas de pré-processamento.

#### Enunciado

Construa um programa que possua como entrada um conjunto de dados Corpus.csv para classificação textual.

#### Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Desenvolver um programa capaz de classificar textos de reviews da Amazon, com 10000 linhas, classificadas em rótulo 1 e 2.
2. Baixe o Notebook para a prática no link disponível na plataforma do IGTI.
3. Fazer o download do dataset “corpus.csv” disponível.

4. Fazer ingestão dos dados no Notebook em que será desenvolvida a prática.
5. Para a execução da atividade, tome como base a aula de Projeto Prático I baseado em técnicas de Machine Learning, apresentado na primeira parte do nosso Módulo, em que apresentamos um mesmo exemplo.
6. Faça as modificações necessárias no Notebook para responder às questões.
7. Para responder às questões de 12 a 15, com base na documentação do Scikilearn e dos algoritmos Naive Bayes e SVM apresentados em nossas aulas, codifique um classificador Random Forest (consulte a documentação do Scikit-learn e tome como exemplo os classificadores Naive Bayes e SVM implementados no Notebook).