



**Data Science
Academy**

www.datascienceacademy.com.br

Deep Learning Frameworks

Mini-Projeto

**Prevendo a Ocorrência de Desastres
Analisando Palavras-Chave em Textos nas
Redes Sociais**



O Twitter se tornou um importante canal de comunicação em tempos de emergência.

A onipresença dos smartphones permite que as pessoas anunciem uma emergência que estão observando em tempo real. Por esse motivo, mais agências estão interessadas em monitorar programaticamente o Twitter (ou seja, organizações de auxílio a desastres e agências de notícias).

Mas nem sempre é claro se as palavras de uma pessoa estão realmente anunciando um desastre.

Este Mini-Projeto mostra um pipeline completo para a tarefa da Processamento de Linguagem Natural de classificação de texto. Especificamente, ele tenta classificar se um tweet descreve um desastre real ou falso.

O pipeline é composto por:

- Limpeza e Pré-Processamento
- Engenharia de Atributos
- Modelo Embeddings de Palavras Pré-Treinado
- Modelo de Deep Learning - Rede Neural Recorrente BiLSTM

Neste Mini-Projeto você estará prevendo se um determinado tweet é sobre um desastre real ou não. Se for desastre a previsão deve ser 1. Caso contrário, 0.

Cada amostra no conjunto de treino e teste tem as seguintes informações:

- O texto de um tweet.
- Uma palavra-chave desse tweet (embora isso possa estar em branco).
- O local de onde o tweet foi enviado (também pode estar em branco).

Ao final do capítulo você encontra o Jupyter Notebook e os datasets. O Jupyter Notebook está comentado linha a linha. Estude, altere o código, experimente e pratique PLN com PyTorch.