

02

Preparando o ambiente

Boas-vindas a mais um curso de Processamento de Linguagem Natural!

Neste curso, vamos criar nossa própria representação **Word2Vec**, para criar um classificador de artigos de jornais, usando apenas os títulos de cada reportagem. Treinado o modelo Word2Vec e o classificador, ambos modelos serão disponibilizados em uma aplicação Web para simular um modelos de machine learning em produção.

Vamos desenvolver nosso curso no [Google Colaboratory](#)

(<https://colab.research.google.com/notebooks/intro.ipynb#recent=true>), você pode ficar livre para desenvolver o curso em um jupyter notebook ou outras plataformas como [Kaggle](#) (<https://www.kaggle.com/notebooks>). Apenas fica o aviso de que o curso vai precisar de um volume considerável de memória.

Para treinar nosso modelo, vamos precisar da base de treinamento e teste. Você pode baixar os arquivos usando os seguintes links:

- [Treinamento](https://caelum-online-public.s3.amazonaws.com/1638-word-embedding/treino.csv) (<https://caelum-online-public.s3.amazonaws.com/1638-word-embedding/treino.csv>).
- [Teste](https://caelum-online-public.s3.amazonaws.com/1638-word-embedding/teste.csv) (<https://caelum-online-public.s3.amazonaws.com/1638-word-embedding/teste.csv>).

Agora, estamos prontos para iniciar nossas atividades.

Bons estudos!