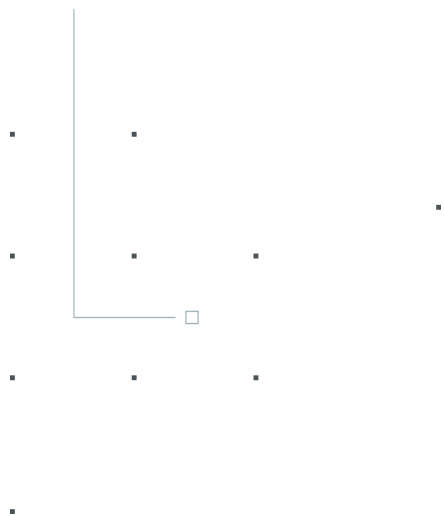


FIAP

NBA



# Generative AI

Prof. Anderson Dourado

- 1. Feature Engineering com Word Embedding**
- 2. BERT e Transformers**
- 3. Demo**
- 4. Exercício**

# Feature Engineering com Word Embedding



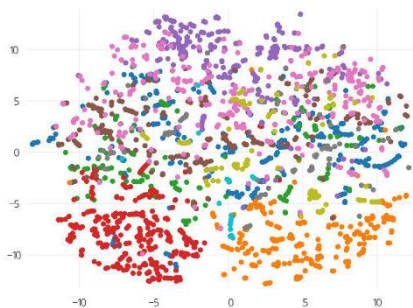
Até aqui vimos como usar **word embeddings generativos (pré-treinados)** focando em análise de similaridade e manuseio do espaço vetorial de alguns algoritmos como Word2Vec.

Existem muitas formas de aplicarmos e utilizamos esses espaços vetoriais, uma delas é fazer um processo de **feature engineering** que envolve **criação e/ou transformação de features para melhorar o desempenho dos modelos de machine learning**. Isso pode incluir a criação de novas features a partir das existentes, normalização de dados, entre outras técnicas.

O objetivo é fornecer ao modelo informações mais relevantes e discriminativas, melhorando assim a capacidade do modelo de aprender e generalizar a partir dos dados.

**A ideia é criar um vetor de features com características de um Embedding generativo e aplicar no nosso problema de classificação de textos.**

Word Embedding pré-treinado



Corpus de documentos rotulados

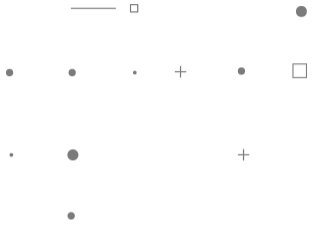


Processo de feature engineering

Vector-space representation

	united	nations	peace
Doc 1	6	9	16
Doc 2	18	13	9
Doc 3	42	17	5
Doc 4	13	11	10

# BERT



- Google BERT é um algoritmo que aumenta a compreensão da linguagem humana;
- Acrônimo de **Bidirectional Encoder Representations from Transformers**;  
(representações de codificador bidirecional de transformadores)
- Permite uma representação contextual das palavras dentro de uma sentença;
- Utiliza a tecnologia de **Transformers**, base dos principais modelos de linguagem utilizados atualmente (GPT, Palm, Llama, RoBERTa, etc...);





O **framework Transformers** em Python é uma biblioteca desenvolvida pela Hugging Face que oferece uma interface de fácil utilização para trabalhar com modelos de **linguagem pré-treinados**.

Como podemos usar o framework Transformers?

- **Fine-Tuning de Modelos de Linguagem:** Você pode carregar modelos pré-treinados usando o framework Transformers, o que economiza muito tempo e recursos em comparação com treinar modelos do zero.
- **Transferência de Aprendizado:** você pode tirar proveito do poder da transferência de aprendizado, aplicando modelos pré-treinados a tarefas específicas de NLP. Isso é útil quando você tem um conjunto de dados limitado e deseja melhorar o desempenho do seu modelo.
- **Recursos de Tokenização e Vetorização:** O Transformers inclui ferramentas de tokenização que ajudam a processar texto em entradas aceitáveis pelos modelos pré-treinados. Isso é importante porque cada modelo tem seus próprios requisitos de tokenização.
- **Suporte a Tarefas Específicas:** Além das tarefas gerais de NLP, como classificação de texto e geração de texto, o Transformers oferece suporte para tarefas mais específicas, como resumo de texto, perguntas e respostas, análise de sentimentos, suporte para classificação e muito mais.



- O mecanismo que permite o modelo identificar que a palavra **ele** está relacionado ao **panetone** e não ao **forno** chama-se *self-attention*, e está intimamente relacionado com a tecnologia de *transformers*.
- Iremos falar mais de transformers...

# Demo e Exercício



# Obrigado!

profanderson.dourado@fiap.com.br



/anderson-dourado

FIAP MBA<sup>+</sup>

Copyright © 2023 | Professor Anderson Vieira Dourado  
Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente proibido sem consentimento formal, por escrito, do professor/autor.

FIAP