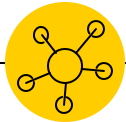


Conversations Toxicity **Detection**



Maestría en Inteligencia Analítica para la Toma de Decisiones
Departamento de Ingeniería Industrial

Lucila Noriega – Andrés Lozano – Julian Baquero



Preprocessing

GloVe: Global Vectors for Word Representation

- Unsupervised learning algorithm for obtaining vector representations for words.
- Training is performed on aggregated global word-word co-occurrence statistics from a corpus
- The resulting representations showcase interesting linear substructures of the word vector space.

<https://nlp.stanford.edu/projects/glove/>



Preprocessing

FastText

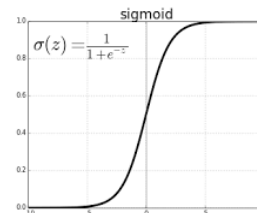
- Open-source, free, lightweight library that allows users to learn text representations and text classifiers
- 300-dimensional pretrained FastText English word vectors released by Facebook research.

<https://fasttext.cc/>

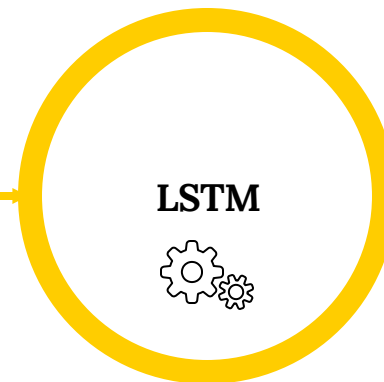
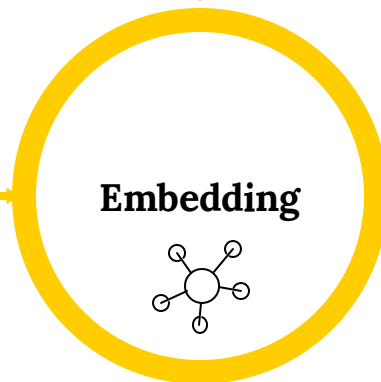
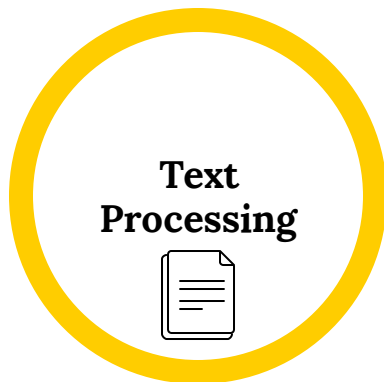
<https://research.fb.com/downloads/fasttext/>



Our process is easy



glove.840B.300d.txt



- ✓ Eliminar símbolos especiales
- ✓ Eliminar contracciones
- ✓ Conversión a minúsculas
- ✓ Eliminar espacios y separadores
- ✓ Eliminación de palabras sin aporte (WordCloud)

crawl-300d-2M.vec

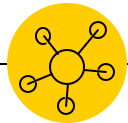


- **Tokenizer**
- **Torch Algorithms**
- **LSTM (Bidirectional)**
- **Tensor**



Referencias

- The Natural Language Processing Group at Stanford University. GloVe: <https://nlp.stanford.edu/projects/glove/>
- Facebook Research. Fast text: <https://fasttext.cc/>, <https://research.fb.com/downloads/fasttext/>



Gracias
