

1

TF-IDF es una técnica ~~de~~ usada en el procesamiento de lenguaje natural para convertir un conjunto de textos en una representación numérica que capture la importancia de las palabras.

El TF significa frecuencia de término y mide cuantas veces aparece un término en un documento.

El IDF significa frecuencia inversa de documentos y mide la rareza de un término.

Esto se puede usar en análisis de temas, filtrado de spam, clasificación de texto, búsqueda de información, etc.

Algunas bibliotecas en python que implementan esto son scikit-learn, spaCy, gensim, nltk, etc.

2) Uno de los problemas más comunes en modelos de N-gramas es la ausencia de n-gramas en el entrenamiento, el laplace smoothing ajusta las probabilidades ~~de~~ al añadir una pequeña cantidad a cada cuenta de palabra o secuencia.

Esto impacta en un modelo de NLP evitando probabilidades cero, distribución de probabilidades más uniforme, mejor generalización, esto hace que el modelo sea más robusto.