Cuestionario clase 47

|  |  |
| --- | --- |
| **la tercera**  CountVectorizer creates a matrix in which each unique word is represented by a column of the matrix | Tercera  TF-IDF es mejor que CountVectorizer porque no sólo se centra en la frecuencia de las palabras presentes en el corpus, sino que también tiene en cuenta su importancia. |
|  |  |
|  |  |
|  | Truncated SVD **generates the matrices with the specified number of columns**, whereas SVD outputs n columns of matrices. It decreases the number of output and better works on the sparse matrices for features output  This estimator supports two algorithms: a fast randomized SVD solver, and a “naive” algorithm that uses ARPACK as an eigensolver on X \* X.T or X.T \* X, whichever is more efficient. |
|  |  |
|  |  |
|  |  |