

1. Santa wants to cluster the stream of letters he receives from children to better assign them to his little helpers. The northpole-datascience-division (NDD) recommends him to use CluStream. What are some possible advantages/disadvantages of this algorithm?



2. Christmas time is approaching fast and Santa is fed up reading all the letters manually. He asks one of his helpers from the northpole-datascience-division (NDD) for ways of getting the most relevant terms from the letters of children to Santa.

Santas little helper recommends...

- A TF-IDF (Term Frequency - Inverse Document Frequency)
- B PCA (Principle Component Analysis)
- C None of above, he shouldn't be so lazy and read the letters manually!