Algorithm:

* Split documents into train and test set
* Cluster training documents using LDA
* For documents in test set:

- Find closet cluster

- Recommend tags using user-based collaborative-filtering or KNN (Distance: Euclidean / Cosine, Weight: TF-IDF)

- Compare recommended tags and real tags