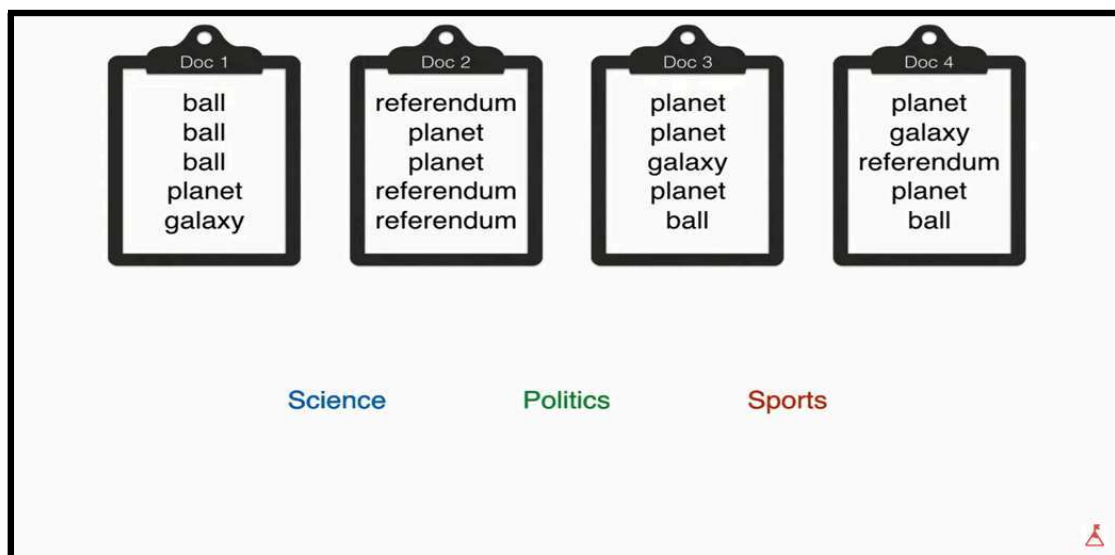


ASSIGNMENT 8 - MINI PROJECT

- **Problem Statement:** Building a personalized news recommender system.
- **Introduction:** News consumption is very subjective in nature. The articles that one prefers to read depends on their personal interests, hobbies and profession. What we have attempted to build through this project is a primitive news recommendation system that sorts random news articles into various fields of interest, and based on the user's choice, recommends other news articles from a similar topic.

Sorting articles into topics:



Consider 4 documents, containing a mixture of 4 words: ball, planet, galaxy and referendum. Using historical knowledge and the theory of association, we humans can classify the documents into the three given categories, based on the frequency of the words used in the document. However, the computer cannot understand English. The documents will look like a compilation of gibberish, the way the following documents seem to us:



All that the computer knows is that the given documents have to be sorted into 3 different topics, each document containing a combination of words that can help indicate the topic of the document.

This is where LDA comes into picture. Latent Dirichlet Allocation is like a machine that is used to sort documents into n topics, based on two parameters - α and β - which are indicative of the importance of the topic to the document and the importance of the word to the topic, respectively.

We have used LDA for topic modeling, and as an alternative, have also presented how KNN classification can be used to sort the news articles into relevant topics.

Subsequently, using cosine similarity, we have recommended relevant news to the user, based on their previous article selection.

- **Dataset Information:**

Link:

<https://www.kaggle.com/kotartemiy/topic-labeled-news-datase>
[t](#)

We have used a subset of 145 entries from this dataset

Snippet of the dataset:

topic	link	domain	published_...	title	lang
SCIENCE	https://www.somagnews.com/spacexs-starship-spacecraft-saw-150-meters-high/	somagnews.com	2020-08-05 17:12:00	SpaceX's Starship spacecraft saw 150 meters high	en
SCIENCE	https://www.sciencealert.com/these-orbs-look-like-candy-but-they-re-actually-different-flavours-of-a...	sciencealert.com	2020-08-13 23:23:31	These Orbs Look Like Candy, But They're Actually Different Flavours of Phobos	en
TECHNOLOGY	https://www.kotaku.com.au/2020/08/come-see-what-its-like-to-play-microsoft-flight-simulator/	kotaku.com.au	2020-08-14 06:29:00	Come See What It's Like To Play Microsoft Flight Simulator	en
TECHNOLOGY	https://www.notebookcheck.net/Xiaomi-rolls-out-Android-10-to-the-Redmi-Note-7-Pro-in-India-as-Mi-Not...	notebookcheck.net	2020-08-11 16:17:38	Xiaomi rolls out Android 10 to the Redmi Note 7 Pro in India as Mi Note 10 and Redmi Note 8 Pro rece...	en

Note by the publishers:

Content

We collected over 100k articles for 8 different news topics

BUSINESS | 15000

ENTERTAINMENT | 15000

HEALTH | 15000

NATION | 15000

SCIENCE | 3774

SPORTS | 15000

TECHNOLOGY | 15000

WORLD | 15000

Those articles got published over the first half of August 2020.