**NETFLIX'S RECOMMENDATION SYSTEM FROM LET’s UPGRADE**

* It is based on rule-based approaches.

**TYPES OF RECOMMENDATION SYSTEM:**

1. content-based.

2. collaborative filtering (based on the hyperparameter).

3. hybrid -based.

**CHALLENGES:**

* cold play to start because the things that we start may not be solved easily and are in different places and far away from each other.
* resources are needed a lot.

**CONTENT-BASED RECOMMEND SYSTEM:**

* It works based on the similarities between two texts or objects.

**for example,**

**FEAR** and **FAIR----** have more similarities.

**F, A, R-----** are similar words.

**FEAR** and **XYWZ** ----have the least similarities, no matches between the words.

* this is how the cosine similarity works between two vectors that are based on the cosine angle

**NLP – Natural Language Processing:**

* It is mainly used to store only important words from a sentence.

**for example,**

* In a movie review **I didn't like pathaan** **because the action scenes were not appealing words**, like I and the and are not important in the sentence only the words not like and not appealing are important in the sentence.
* The words that are not important are known as garbage words or values.
* They are called stop words so when we clean data, we remove all these stop words.
* So that when we use NLP for the text version we just use it to remove the stock words so that our model does not need to contribute to the words and waste its time and to improve our model efficiently.

**RE-REGULAR EXPRESSION:**

* Making or building rules that you want to remove from your dataset.

**STEMMING**:

* It means it removes the suffixes and prefixes from the words.

**for example,**

**runner** will be cut into **run.**

**Er** will be cut off.

* **[ed, ing, er, ining, ly]** are words not needed and will be cut off and whatever is left is known as the needed word from the sentence.
* It is different from **LIMITATION** because it needs a dictionary that is like in the dictionary form it has a word with an adjective, noun, and adverb. So errors are less. but in stemmer, the word may be cut into runner can be runn which is wrong but it is mainly used because it is light weighted.

**TfidfVectorizer:**

* It is used to convert the dataset into vector form so that the number of occurrences of the word in the sentence is calculated and compared with the entire dataset.
* This will help us to find the importance of the word in the sentence but also in the entire dataset that we have and find how common or uncommon the word is in the given dataset.

**COSINE SIMILARITY:**

* It is used to map the distance of the same words in the dataset so that it is easy to group all of them.