Natural Language Processing and Unsupervised Learning Project - Song Recommendation

Calvin Yu

Motivation

- Interested to understand the concept of Spotify song recommendation system
- Like Music !

Data

- Song Lyrics Dataset(Kaggle)

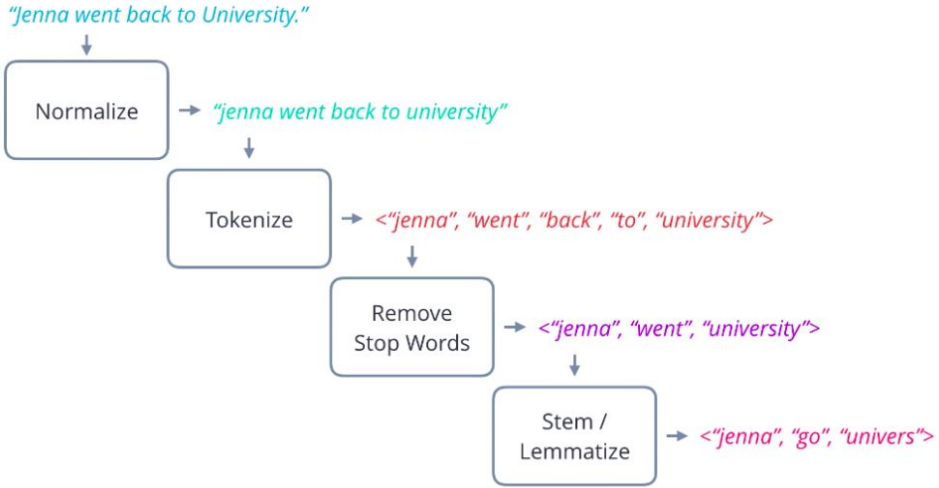
  Male:Drake, Edsheeran, Eminem, Justin Bieber , Maroon 5

  Female :Ariana Grande, Beyonce,Dua Lipa, Lady Gaga, Taylor Swift

- 1526 rows with 8 features (mainly focusing on Lyrics)

- EDA(drop duplicates,missing values)

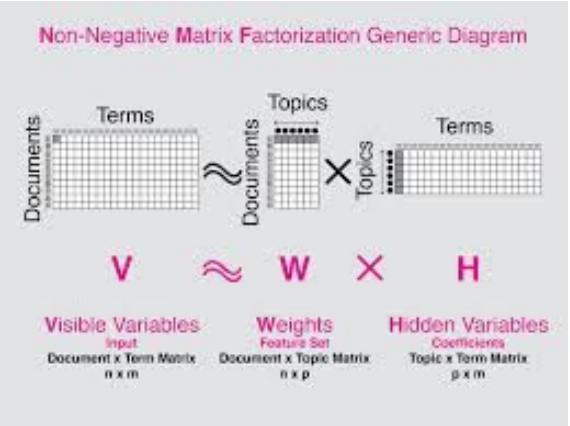
Word pre-processing



- Create a Lemmatization tokenizer
- Create a stopword list and extend it with meaningless words
- Fit them to the tfidf Vectorizer (fit-transform)

Text Analysis

- Non-Negative Matrix Factorization (NMF)
- Latent Dirichlet Allocation (LDA)
- Correlation Explanation (CoRex)



Topic1 - Rap

Topic 2 - Falling in Love

Topic 3 - Girl

girl

think

lonely

time

dance

crazy

work

Jok

world

Topic 4 - Missing someone

home
think
time
see
feel
say
night
could
way
everything

Topic 5 - Confessing

want

light
break darling
romance free
lady bad

body

Topic 6 - Needy

Topic 7- Regret

Topic 8 - Breaking up

Song Recommendation

- Each song has its unique topic score on each topic
- Apply Cosine Similarity on the topic scores
- Will give me songs that are similar to each other (based on lyrics)

Song Recommendation

```
recommendation(lyric_matrix,song,cos_sim=cos_sim)
```

```
Please input a song: perfect
```

```
: recommendation(lyric_matrix,song,cos_sim=cos_sim)
```

```
Please input a song: perfect
1 ['americano' 'Lady Gaga']
2 ['bloody mary' 'Lady Gaga']
3 ['summer games' 'Drake']
4 ['love yourself' 'Justin Bieber']
5 ['teenage fever' 'Drake']
6 ['born this way - zedd remix' 'Lady Gaga']
7 ['up' 'Justin Bieber']
```

Further steps

- Add more songs
- Put song tempo and genre into the cosine similarity metric
- Try different numbers of components