

---

# THE GLOBALIZING K-POP PROJECT: SENTIMENT ANALYSIS OF GLOBAL FANDOMS ON SOCIAL MEDIA

**Supervised by:**

Dr. Stacey Scott

Dr. Ritu Chaturvedi

**Team Members:**

1. Aditi Satsangi ‘
2. Krish Garg

<sup>1</sup> Presented by: Gunpreet Singh



# Contents

---



- 1 Objective
- 2 Introduction
- 3 Background
- 4 Research Workflow
- 5 Methodology
- 6 Results
- 7 Evaluation and Comparative Study
- 8 Conclusions
- 9 References

# Objective

Find the evidence of social supports like emotional, informational, appraisal, etc using K-pop data obtained from social media platforms.



# Introduction



- The Globalizing K-pop project uses social media and machine translation to:
  - Enhance the well-being and sense of community among global K-pop fans
  - Improve access to Korean K-pop media
- Fandoms are communities with shared interests in media or celebrities that transform and curate content
- Global fandoms rely on social media to:
  - Connect fans from diverse geographic regions, cultures, languages, and abilities

# Background

## What is K-pop?

K-pop, short for Korean popular music, is a form of popular music originating in South Korea as part of South Korean culture.

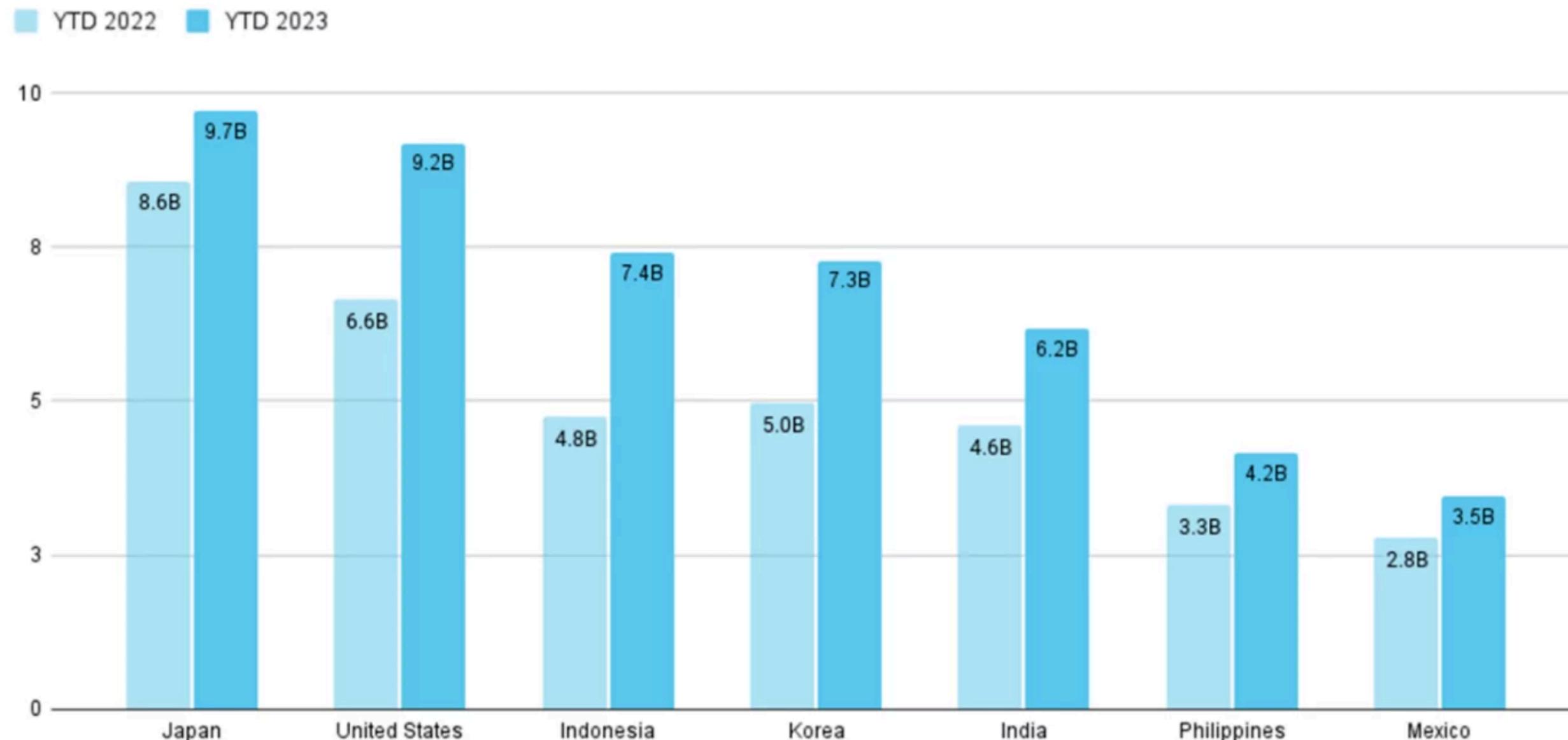
<https://en.wikipedia.org/wiki/K-pop>



# Why K-pop?

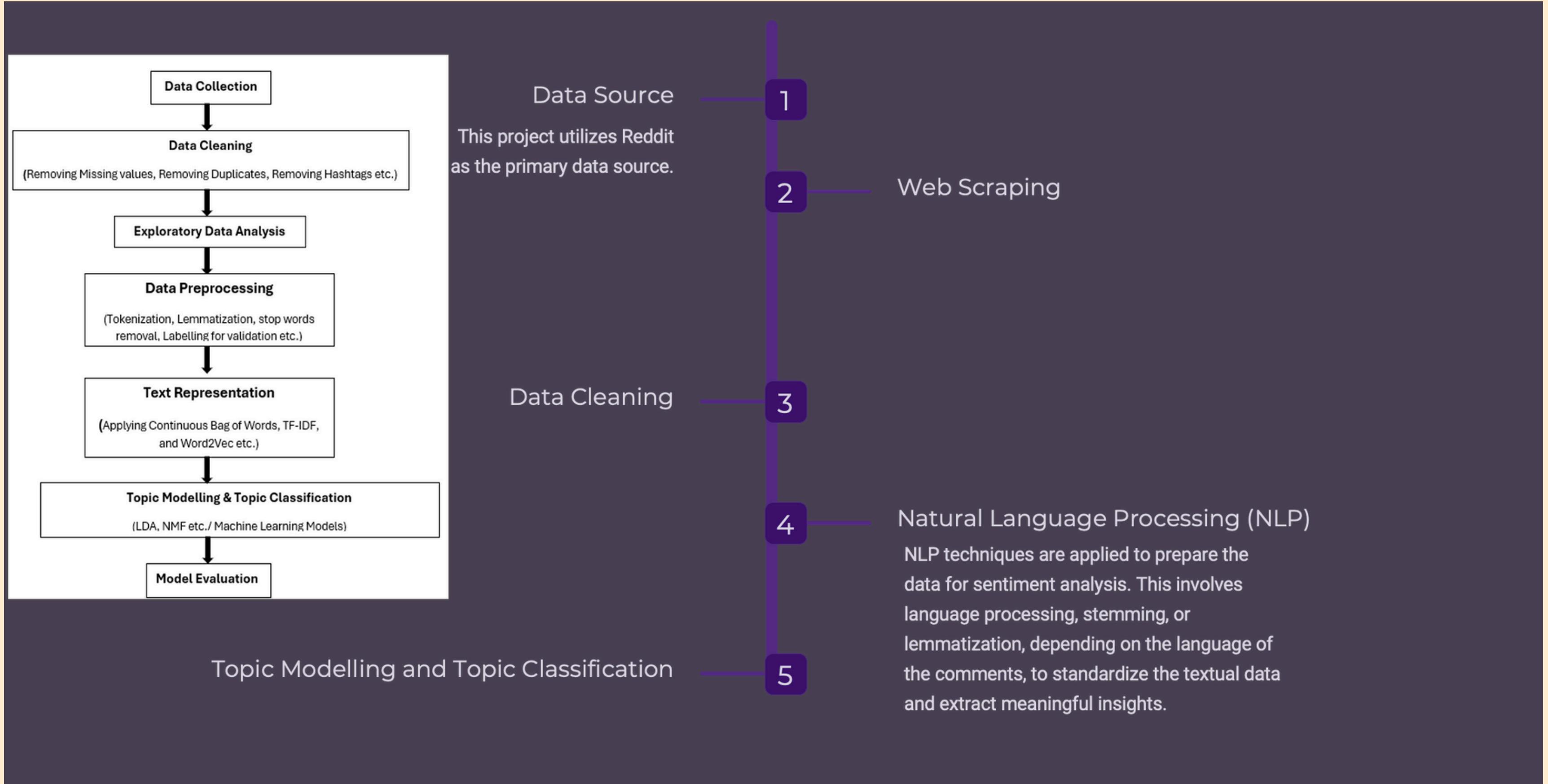
## Streaming of Top 100 K-Pop Artists by Country

In Billions, Total Streaming On-Demand, Year-to-Date as of Week Ending 10/5/23



<https://luminatedata.com/blog/mapping-out-k-pops-global-dominance/#:~:text=Looking%20at%20the%20Top%20100,through%20the%20same%20time%20period.>

# Research Workflow



# METHODOLOGY

**Data Collection and Preprocessing:** Load and preprocess the text data.

1

**Text Representation:** Convert text data into numerical form using TF-IDF and voyage-AI

2

**Model Training:** Train using Models like LDA, NMF, LSA and BERTopic

3

**Visualization of Clusters:** Using Word Cloud and Coherence Score

4

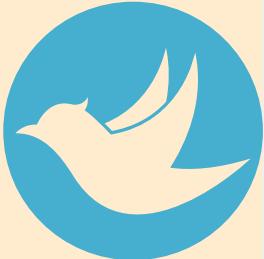
**Top Keywords Extraction:** Extract and visualize top keywords for each topic.

5

**Representation Model:** Using models like MMR, KeyBERT and Llama2 for representation

6

# DATA COLLECTION



## TWITTER DATA

THE TWITTER DATA FOR EACH FANDOM WAS EXTRACTED USING THE TWITTER API IN LAST SUMMER 2023

	conversation_id	text	reply_count	first_tweet	text	cleaned_text
0	1658489916412801024	Fantasy friends plz help me to find the nicest...	2	Fantasy friends plz help me to find the nicest...	['fantasy', 'friend', 'plz', 'help', 'find', '...	fantasy friend plz help find nicest full body ...
1	1658831414291226624	praying that kflex london has an amazing lineu...	1	praying that kflex london has an amazing lineu...	['praying', 'kflex', 'london', 'amazing', 'lin...	praying kflex london amazing lineup just like ...
2	1658872268708265984	im already tired but still need to stay up to ...	1	im already tired but still need to stay up to ...	['already', 'tired', 'still', 'need', 'stay', ...	already tired still need stay assignment help ...
3	1658916239949930553	I will probably delete this tweet later but I ...	1	I will probably delete this tweet later but I ...	['will', 'probably', 'delete', 'tweet', 'later...	will probably delete tweet later just got inte...
4	1659090060279529472	everyone pray for my accounting tmrw @i520rick...	5	everyone pray for my accounting tmrw	['everyone', 'pray', 'accounting', 'tmrw', 'aw...	everyone pray accounting tmrw aww thank bff ri...

Data Size : 979



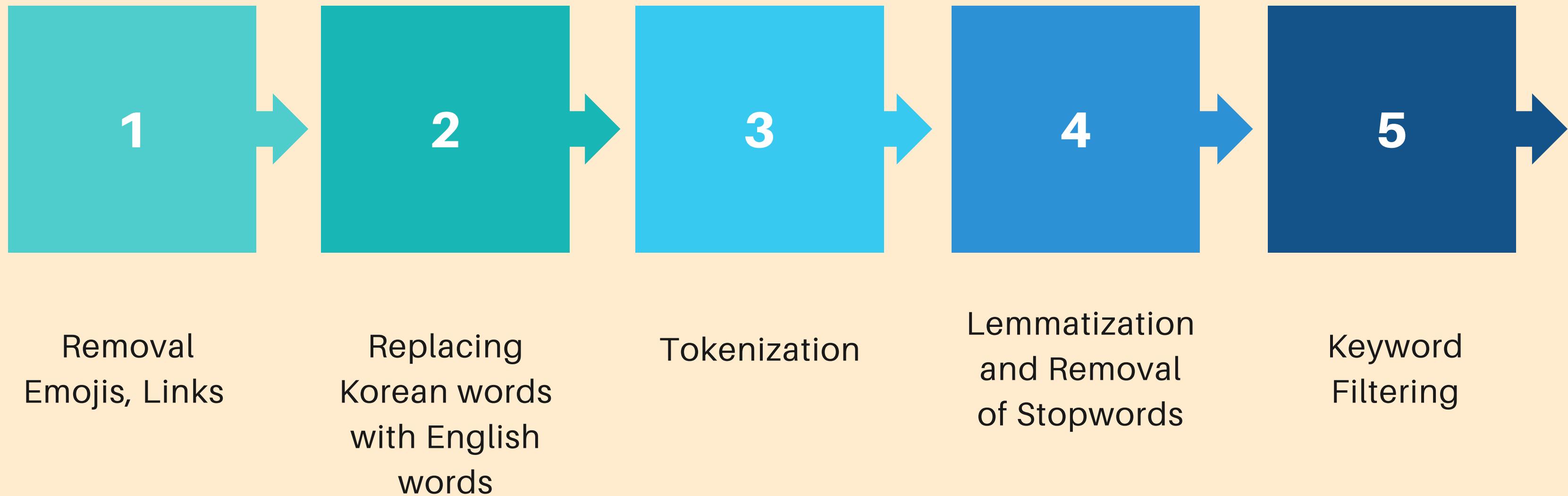
## REDDIT DATA

THE REDDIT DATA FOR EACH K-POP FANDOM WAS EXTRACTED USING THE PRAW LIBRARY

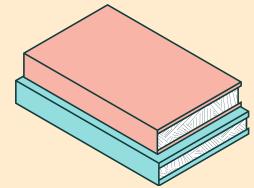
reddit	Title	Body	URL	Comments	Emotional Support Instances in Title	Emotional Support Instances in Body	Emotional Support Instances in Comments	combined_text	cleaned_combine
twice	240610 Weekly Discussion Thread	Once!\\n\\nWelcome to our weekly discussion ... Hey	https://www.reddit.com/r/twice/comments/1dc82s...	MapleLeafsFan3: Anyone ever just like...sit do...	0	0	0	['anyone ever just like...sit do...']	240610 Weekly Discussion Thread MapleLeafsFan3...
twice	240614 YouTube Update - NAYEON "ABCD" M/V	NaN	https://www.youtube.com/watch?v=oUZtxRcPZw	veritek25: Nayeon that was awesome - Na surely...	0	0	0	['stayed up til midnight just to watch this e...']	240614 YouTube Update - NAYEON "ABCD" M/V veritek25: Nayeon that was awesome - Na surely...

Data size: 3396

# DATA PREPROCESSING

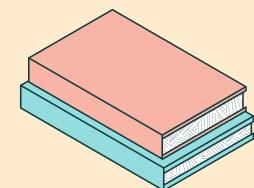


# Text Representation: Mapping Words to Meaningful Vectors



## TF-IDF

The Term Frequency-Inverse Document Frequency (TF-IDF) technique is used to calculate the importance of words within the dataset. This weighting scheme considers the frequency of words in each comment and their overall presence across the entire dataset. Words that appear frequently in a particular comment but are relatively rare across the corpus are deemed more significant.



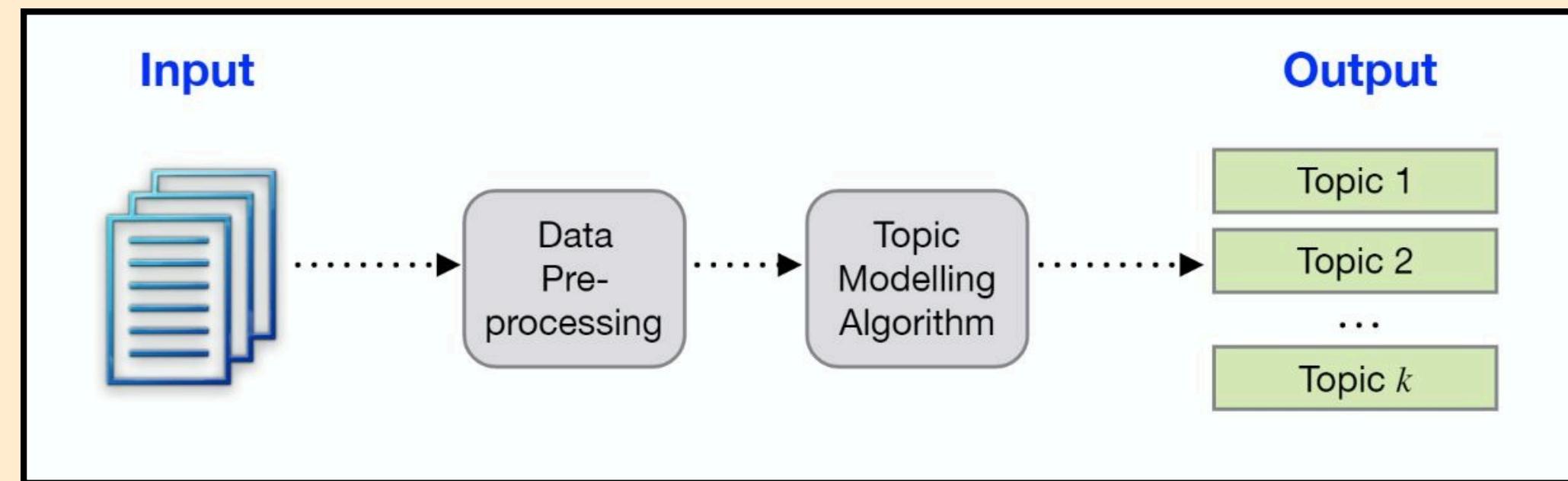
## Voyage-AI

Instruction-tuned general-purpose embedding model optimized for clustering, classification, and retrieval.

# What is Topic Modelling?

In natural language processing (NLP), topic modeling is a text mining technique that applies unsupervised learning on large sets of texts to produce a summary set of terms derived from those documents that represent the collection's overall primary set of topics.[1]

[1] [https://www.ibm.com/topics/topic-modeling#:~:text=In%20natural%20language%20processing%20\(NLP,overall%20primary%20set%20of%20topics.](https://www.ibm.com/topics/topic-modeling#:~:text=In%20natural%20language%20processing%20(NLP,overall%20primary%20set%20of%20topics.)



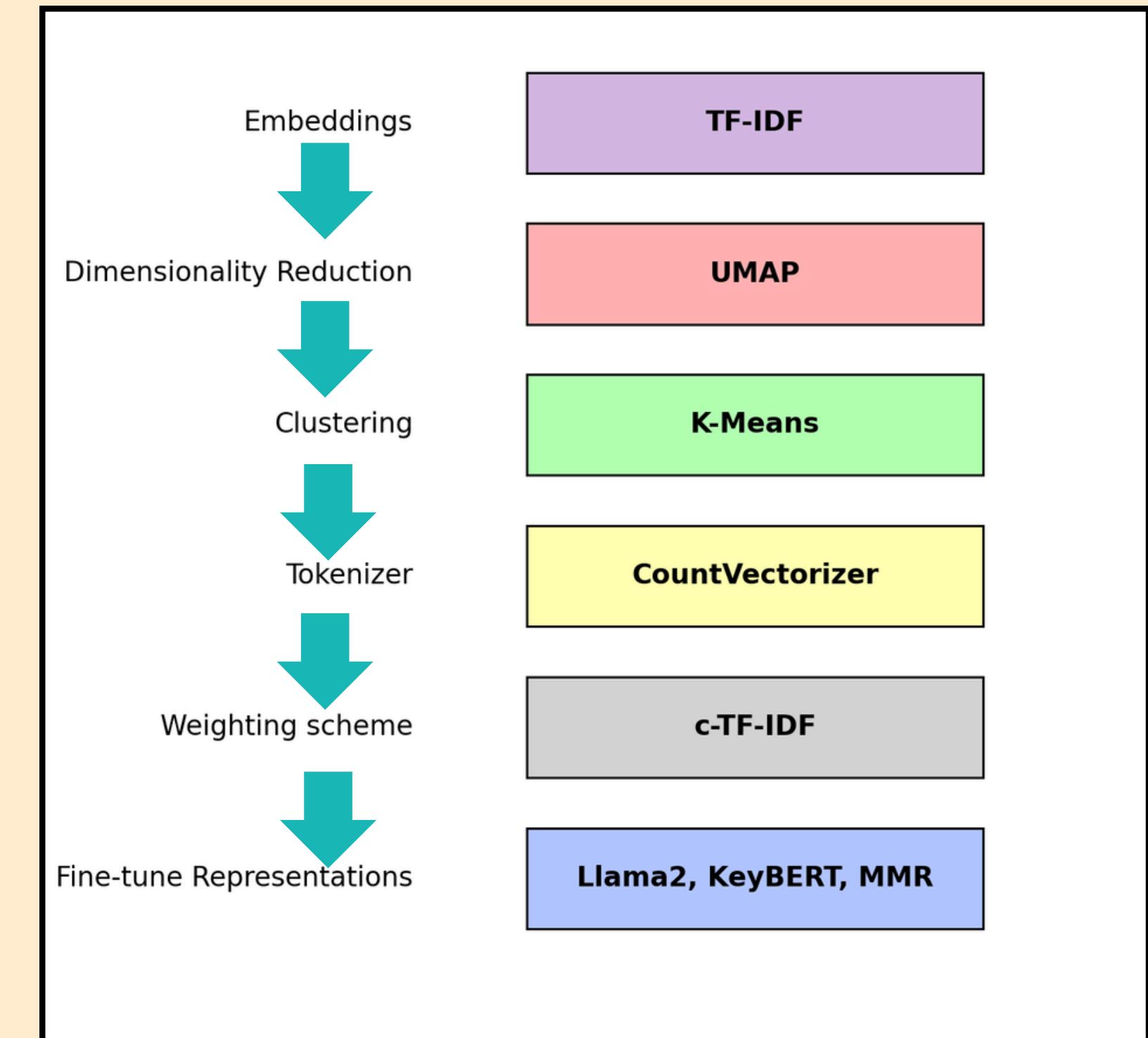
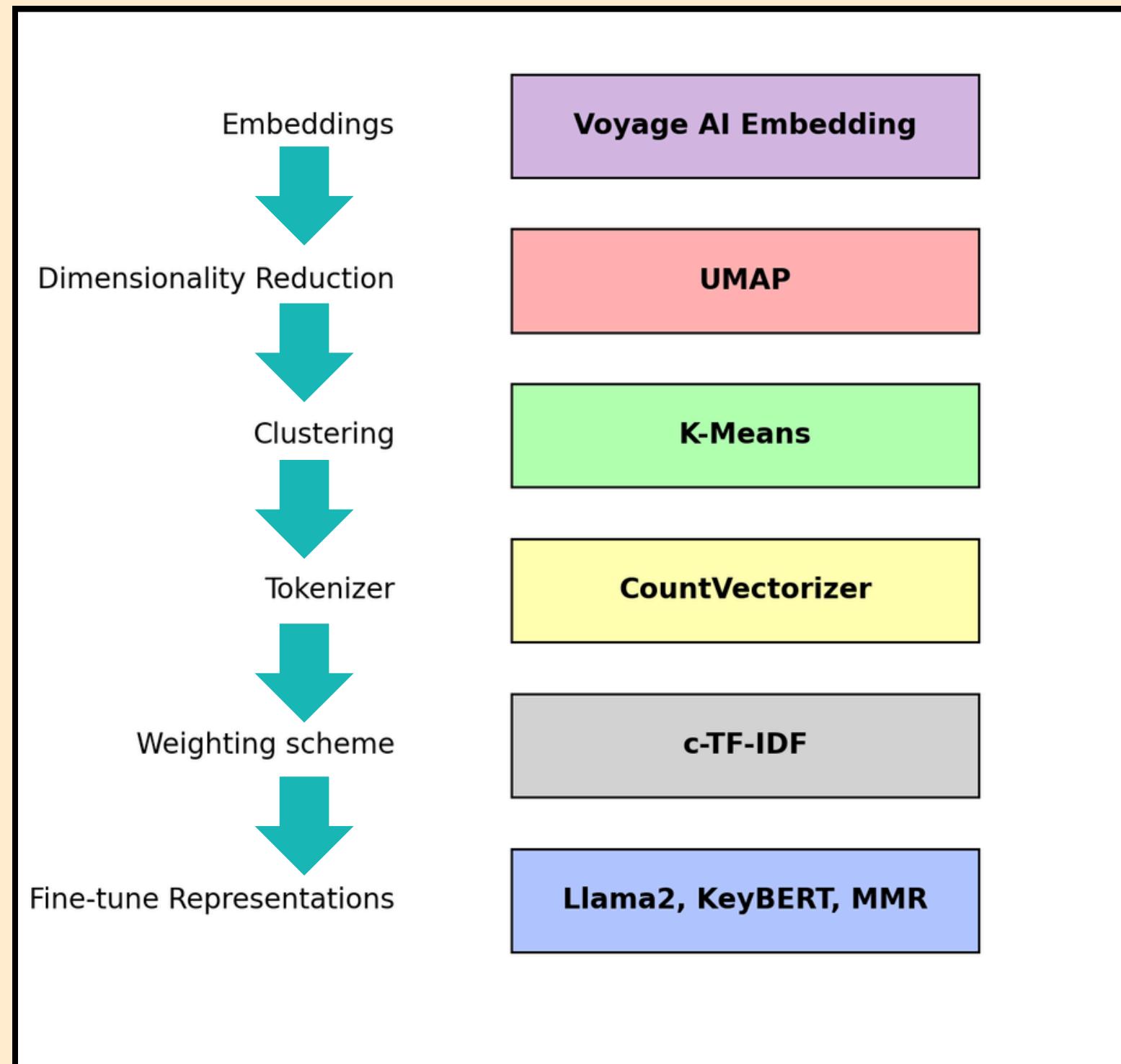
<https://hackernoon.com/nlp-tutorial-topic-modeling-in-python-with-bertopic-372w35l9>

## BERTopic

BERTopic is a topic modeling technique that leverages Huggingface 😊 transformers and c-TF-IDF to create dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions.[2]

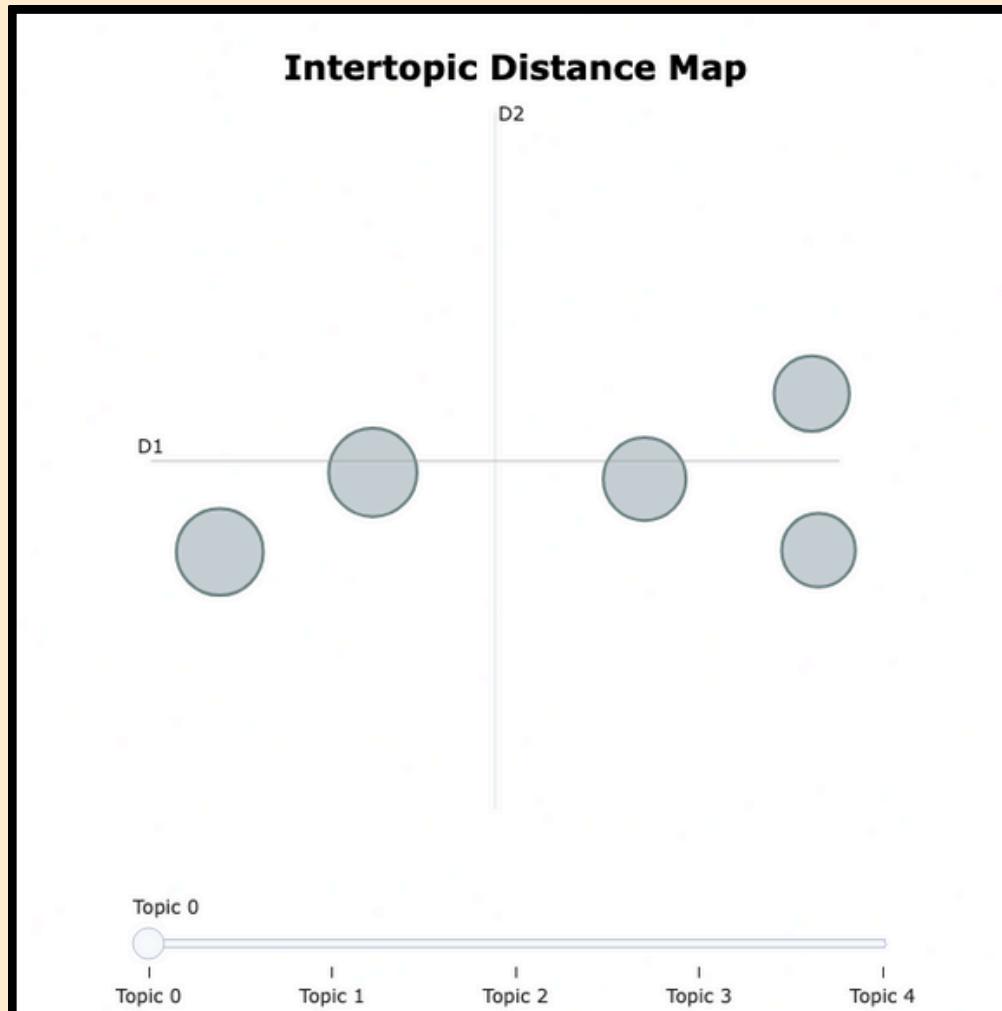
[2] <https://maartengr.github.io/BERTopic/index.html>

# BERTopic Architectures



# Results

## Comparison of different Topic Models- Coherence Scores



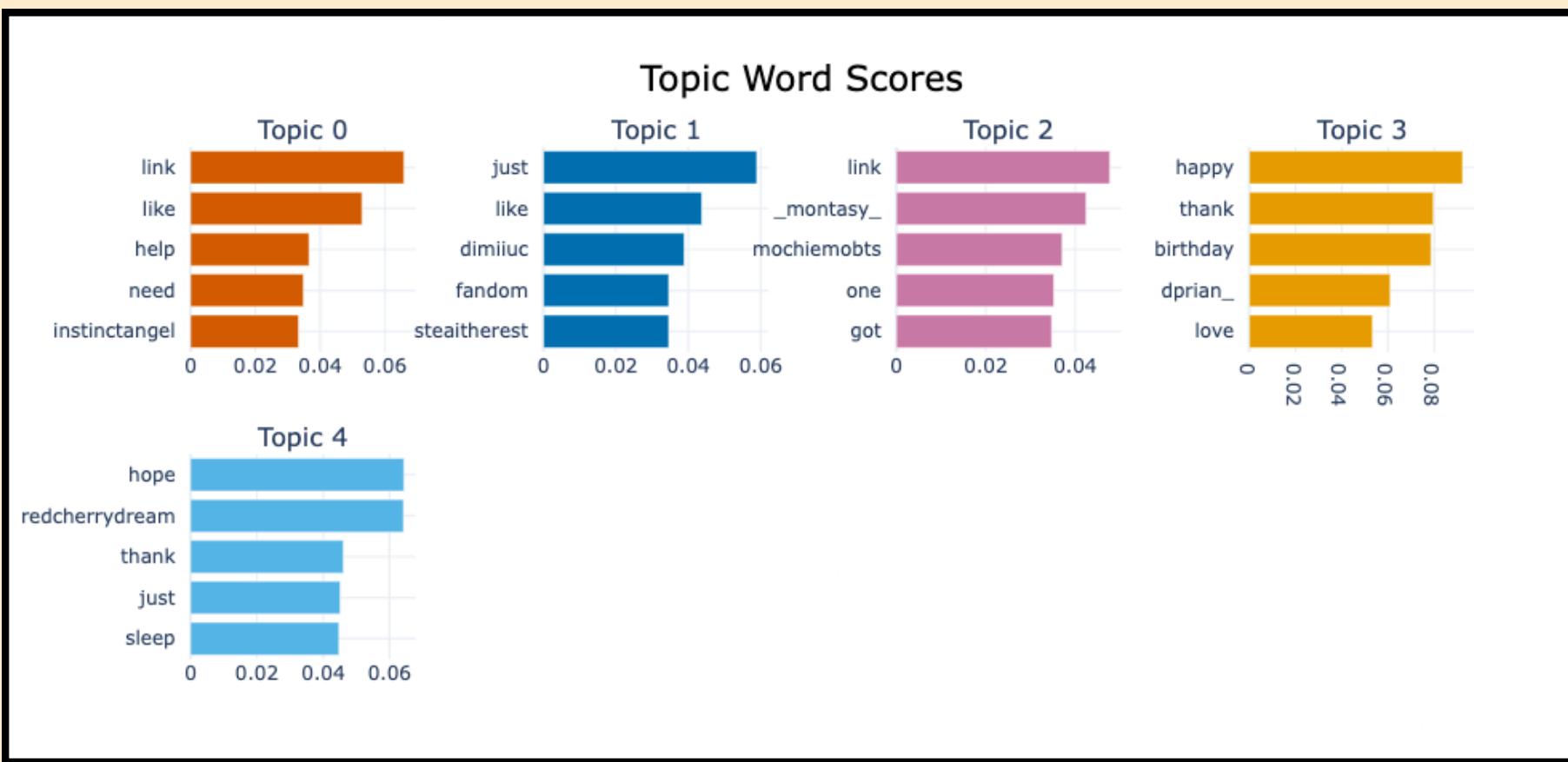
### BERTopic

Metrics	Voyage AI	TF_IDF
U_mass	-5.492	-3.104
CV	0.588	0.444

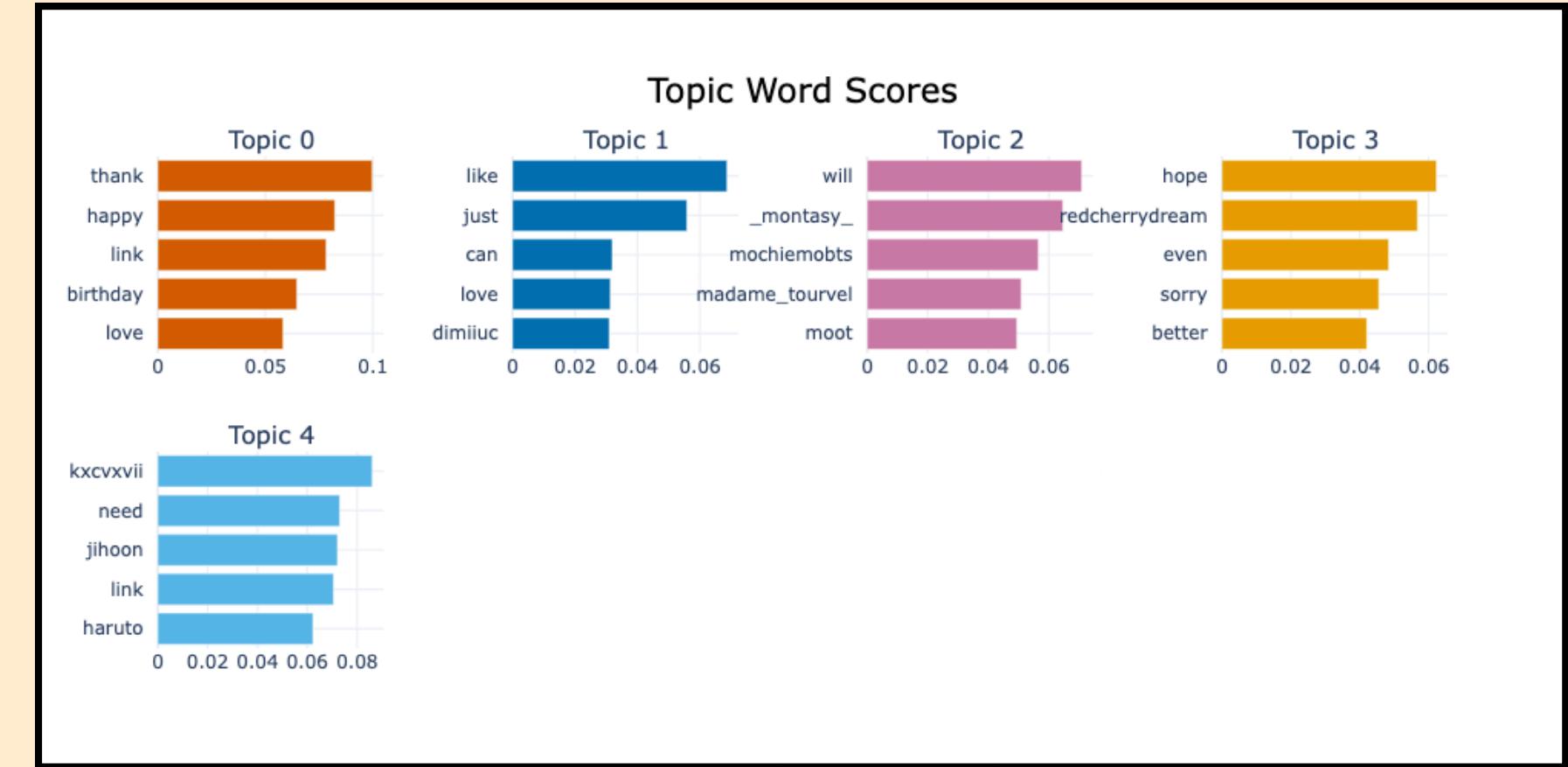
Topic Model	Coherence Score
LDA	0.634
NMF	0.54
LSA	0.368

# Top Keywords - Bar Chart Representation

## Voyage-AI



## TF-IDF



# Representation Models

## Voyage-AI

### Llama2

<b>Topic 0:</b>	'Mental Health and the Use of Packing Tape'
<b>Topic 1:</b>	'Fandom Discussions and Personal Experiences'
<b>Topic 2:</b>	"K-pop and K-beauty enthusiasts' online interactions and preferences"
<b>Topic 3:</b>	'Happy Birthday'
<b>Topic 4:</b>	'Mental Health and Sleep Deprivation'

### KeyBERT

'stuff'	'lf'	'smt'	'mean'	'like'
'fandom'	'feel'	'life'	'whatever'	'keep'
'yooniquelyja'	'krklemm'	'kebinwooo'	'__ysyh'	'원어스'
'birthday'	'day'	'happy'	'congratulation'	'life'
'rest'	'break'	'sleep'	'feeling'	'hope'

### TF-IDF

### Llama2

<b>Topic 0:</b>	'Happy Birthday'
<b>Topic 1:</b>	'Social Media Discourse and Online Fandom'
<b>Topic 2:</b>	'K-pop fandom and merchandise'
<b>Topic 3:</b>	'Mental Health and Self-Care'
<b>Topic 4:</b>	'K-pop and social media culture'

### KeyBERT

'birthday'	'day'	'congratulation'	'happy'	'thankful'
'fandom'	'tattoo'	'mean'	'feel'	'say'
'yooniquelyjae'	'tweet'	'soobinurhome'	'yootae'	'spcygomtok ki'
'rest'	'break'	'work'	'sleep'	'feeling'
'kxcvxxviii'	'handsome'	'tomorrowbyskz'	'yall'	'jihoon'

# Representation Models

LSA

	Keybert Keywords	GPT	BART
0	[love, time, need, thank, help]	Emotional Support Requests and Responses	someone send cute clip vids photo gifs faves p...
1	[concert, birthday, jaengpup, happy, cute]	Gratitude and Positive Celebrations	got birthday balloon believe turning old hag n...
2	[price, buy, homework, heart, rt]	Help and Assistance Requests	help link gunwookeg help decide live wallpaper...
3	[twowomenkpop, dahlihw, song, morning, sleep]	Greetings and Daily Interactions	good morning people phone ily link dahlihw go...
4	[jaengpup, song, dahlihw, morning, heart]	Music and K-Pop Discussions	good morning mooties feel star proud boy dropp...

NMF

	Keybert Keywords	GPT	BART
0	[redcherrydream, hope, going, better, need]	Emotional Support and Encouragement	wanted studying think cancel now arm started a...
1	[birthday, congratulation, twowomenkpop, congr...	Celebrations and Well-Wishes	got birthday balloon believe turning old hag n...
2	[pay, rt, buy, moot, homework]	Requests for Help and Opinions	help link gunwookeg hi need declutter na free ...
3	[dahlihw, morning, feeling, song, link]	Morning Greetings and Positive Vibes	good morning people phone ily link dahlihw go...
4	[jaengpup, heart, love, miss, dimiiuc]	Request for Assistance and Community Support	really rough day mentally just want thank gave...

LDA

	Keybert Keywords	GPT	BART
0	[support, congratulation, bday, look, face]	Fandom Engagement and Conversations	worry just
1	[moot, cost, live, congratulation, stay]	Technical Help and Problem-Solving	wthhh
2	[envelope, hair, happiness, era, congratulation]	Expressions of Affection and Gratitude	help link gun
3	[cost, hair, gon, online, congratulation]	Health, Wellness, and Emotional Support	taken weekend
4	[leehan, congratulation, twitter, genuinely, k...]	Concerts, Events, and Fandom Activities	help rt

# Future Plan

- Collect twitter or reddit data with replies to the tweet or the main reddit.
- Do the external validation of the results.
- Try the Llama 3.1 model for the representation of the clusters.
- Hyperparameter tuning of BERTopic according to the data available in the future.

[https://www.google.com/search?q=k+pop+anime&sca\\_esv=e2fa7d0179107c0a&sca\\_upv=1&udm=2&biw=1440&bih=813&sxsrf=ADLYWIIaQ1tCmcZJ3pnMAXY8ArkI8yIOrg%3A1722954809505&ei=OTSyZs\\_HHqm40PEP\\_d-mwAc&qg=k+pop+a&gs\\_lp=Egxnd3Mtd2I6LNlcAiB2sgcG9wlGEqAggAMgUQABiABDIFEAAYgAQyBRAAGIAEMgUQABiABDIFEAAYgAQyBRAAGIAEMgUQABiABDIFEAAYgAQyBRAAGIAEMgUQABiABEj8GICSCFiaEHACeACQAQCYAVSgAaYBqgEBMrgBACgBApgBAZgCBKACvwHCAhAQABiABBixAxhDGIMBGloFwgIEEAAYA8ICChAAGIAEGEMYigWYAwCIBgGSBwE0oAe3CQ&sclient=gws-wiz-serp#vhid=9FqUmljtdVcNM&vssid=mosaic](https://www.google.com/search?q=k+pop+anime&sca_esv=e2fa7d0179107c0a&sca_upv=1&udm=2&biw=1440&bih=813&sxsrf=ADLYWIIaQ1tCmcZJ3pnMAXY8ArkI8yIOrg%3A1722954809505&ei=OTSyZs_HHqm40PEP_d-mwAc&qg=k+pop+a&gs_lp=Egxnd3Mtd2I6LNlcAiB2sgcG9wlGEqAggAMgUQABiABDIFEAAYgAQyBRAAGIAEMgUQABiABDIFEAAYgAQyBRAAGIAEMgUQABiABDIFEAAYgAQyBRAAGIAEMgUQABiABEj8GICSCFiaEHACeACQAQCYAVSgAaYBqgEBMrgBACgBApgBAZgCBKACvwHCAhAQABiABBixAxhDGIMBGloFwgIEEAAYA8ICChAAGIAEGEMYigWYAwCIBgGSBwE0oAe3CQ&sclient=gws-wiz-serp#vhid=9FqUmljtdVcNM&vssid=mosaic)

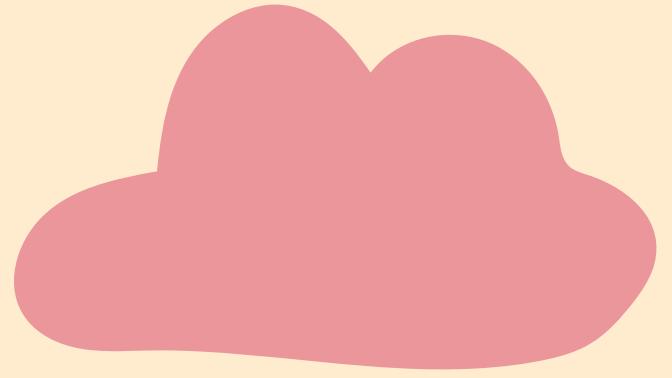
# Conclusion

---

The Globalizing K-pop project has effectively used social media and machine translation to analyze the sentiments and community well-being among global K-pop fandoms. By examining data from Twitter and Reddit with advanced topic modeling techniques like BERTopic, the study identified key themes such as emotional investment, mental health, good wishes, personal reflections, and fan culture. This research underscores the role of social media in connecting diverse fans worldwide. Future efforts will focus on further data collection, validating results, and utilizing advanced models. Overall, this project enhances understanding of the global K-pop phenomenon and aims to improve accessibility and inclusivity of K-pop media.

# References

- 1) Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." arXiv preprint arXiv:2203.0579 4 (2022).
- 2) Lee, K., Toutanova, K., and Devlin, J. "Topic Modeling Algorithms and Applications: A Survey." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 4171-4186. <http://journal.uob.edu.bh>.
- 3) Mostafazadeh, N., et al. "A Review on Sentiment Analysis from Social Media Platforms." Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 908-918. <http://journal.uob.edu.bh>.
- 4) Kumar, B.V Pranay, and Dr. Manchala Sadanandam. "A Fusion Architecture of BERT and RoBERTa for Enhanced Performance of Sentiment Analysis of Social Media Platforms." International Journal of Computing and Digital Systems, vol. 15, no. 1, Jan. 2024, <http://dx.doi.org/10.12785/ijcds/150105>.
- 5) <https://huggingface.co/spaces/mteb/leaderboard>
- 6)<https://llama.meta.com/docs/get-started/>



# THANKYOU

Any Questions ?

