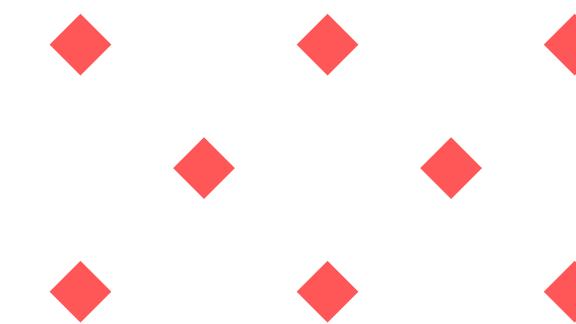
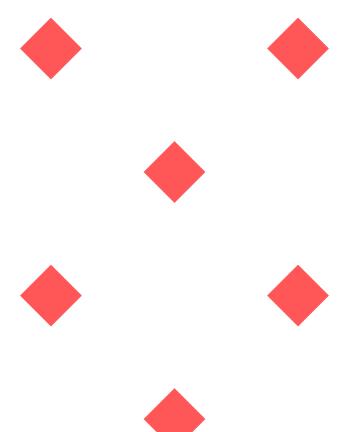


TEAM 31

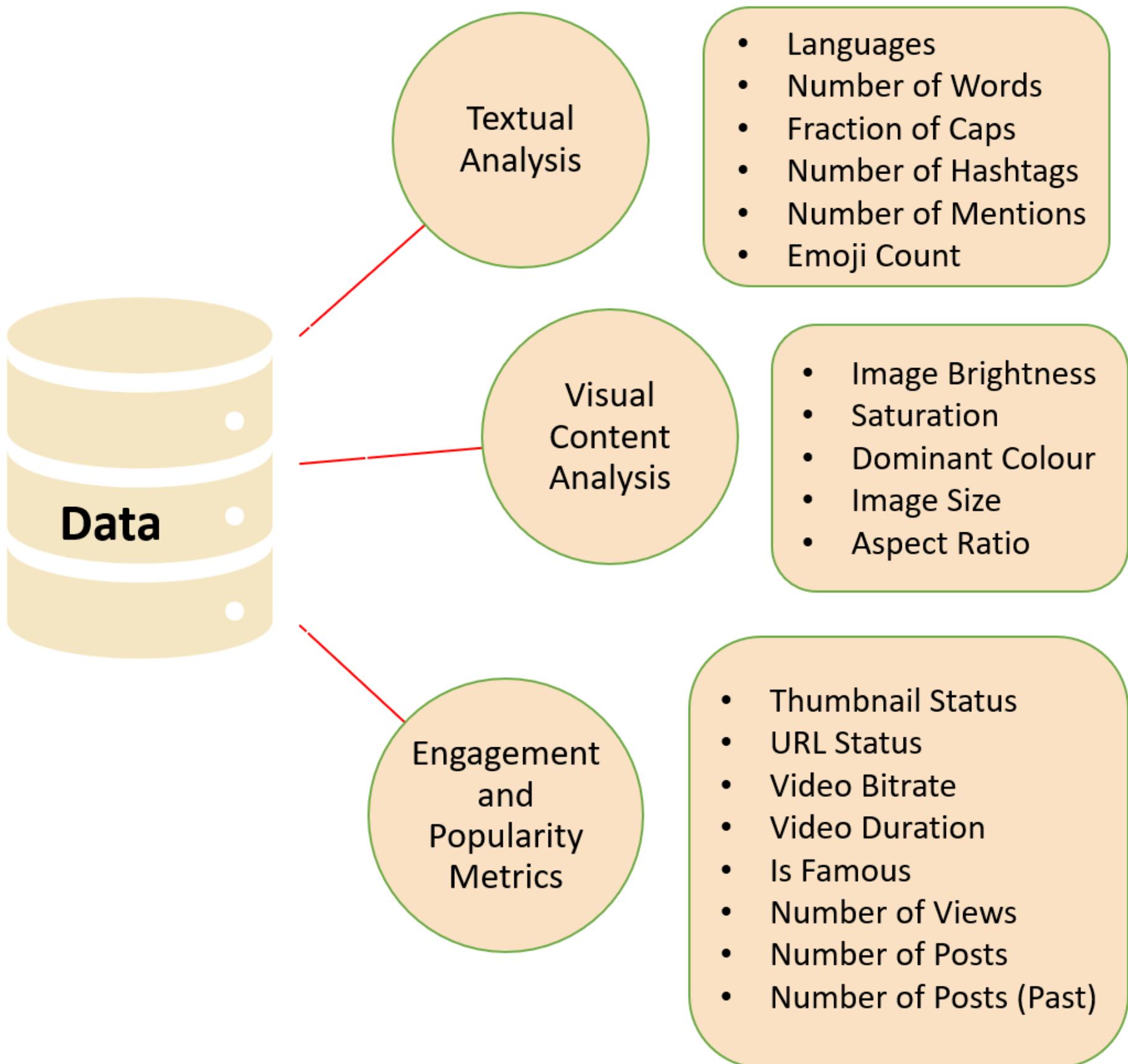


# BEHAVIOUR SIMULATION CHALLENGE

Mid Prep Challenge



# Feature Engineering

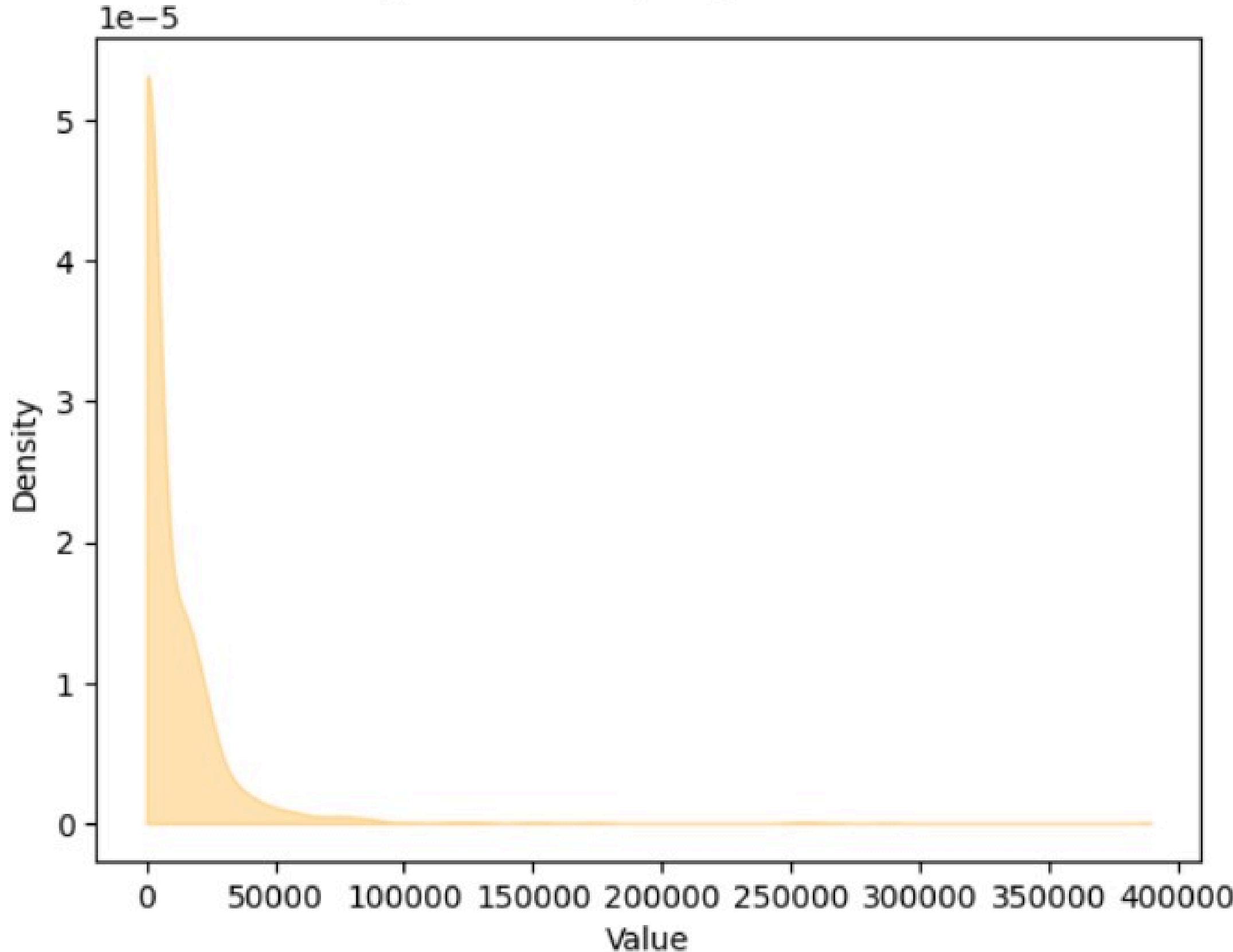


## Username Processing

- High ranking usernames - distinct one-hot encoding
- Low ranked ones share the same one-hot encoding

# Data Imbalance

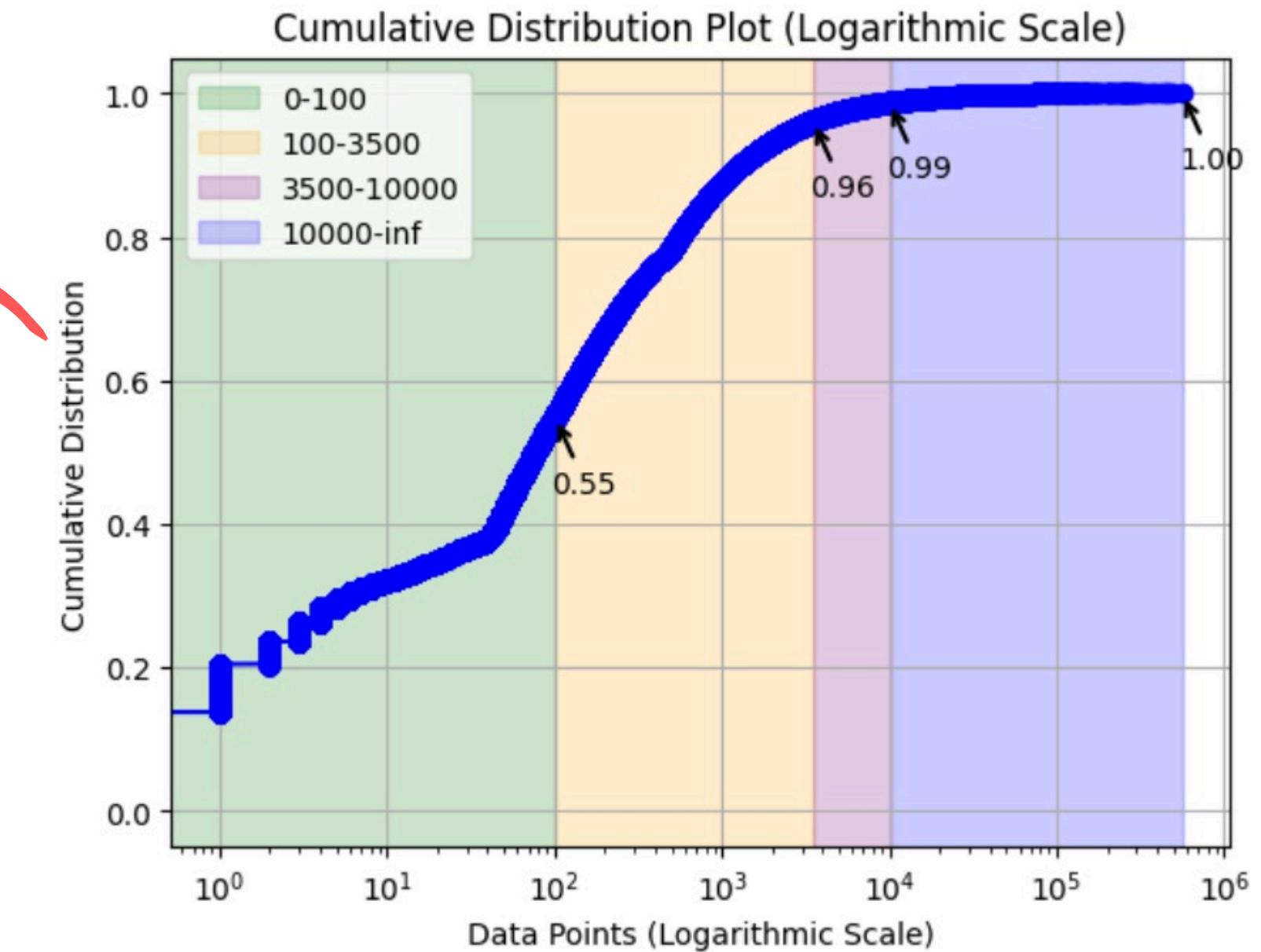
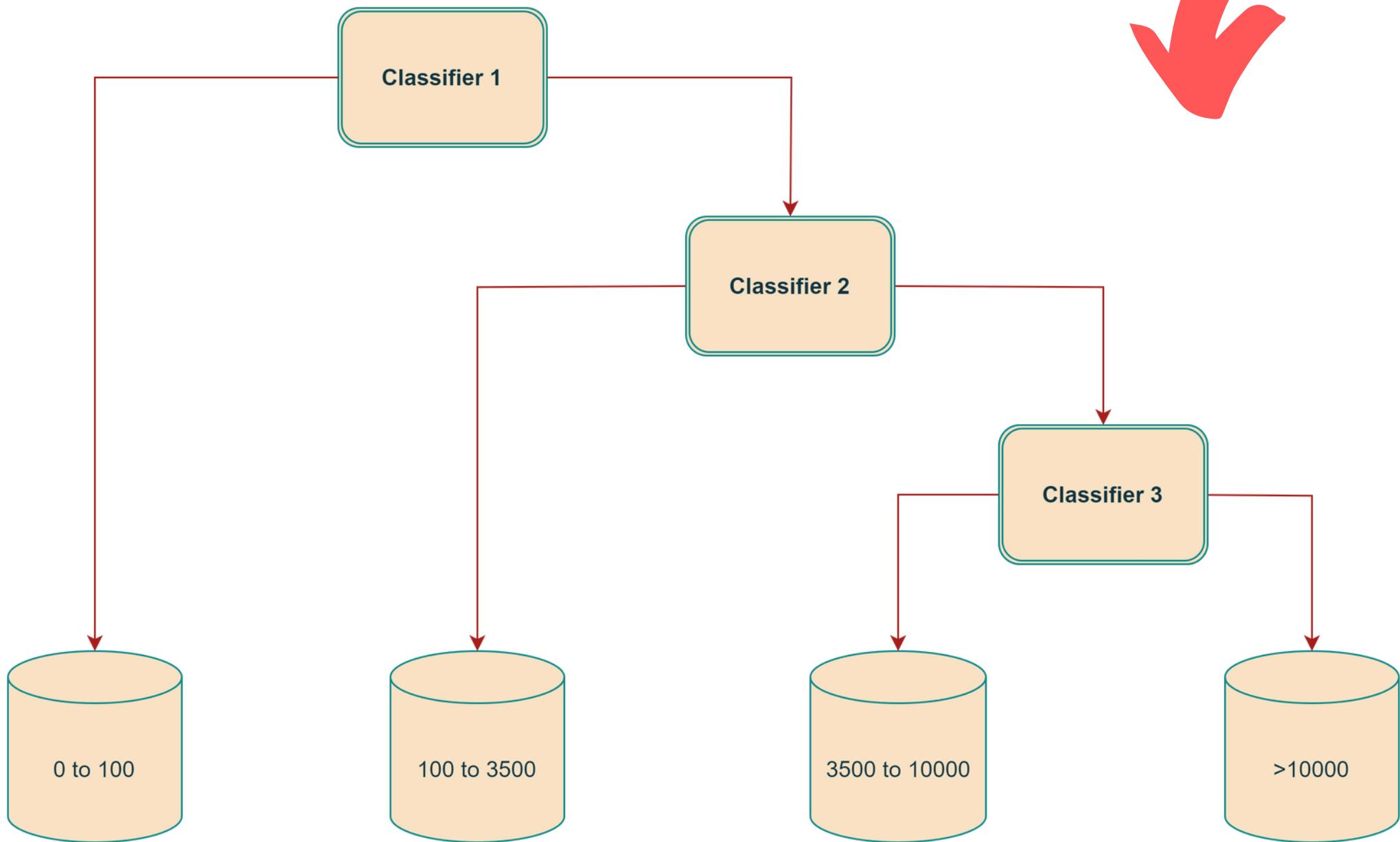
Kernel Density Estimation (KDE) Plot with Mass Marked



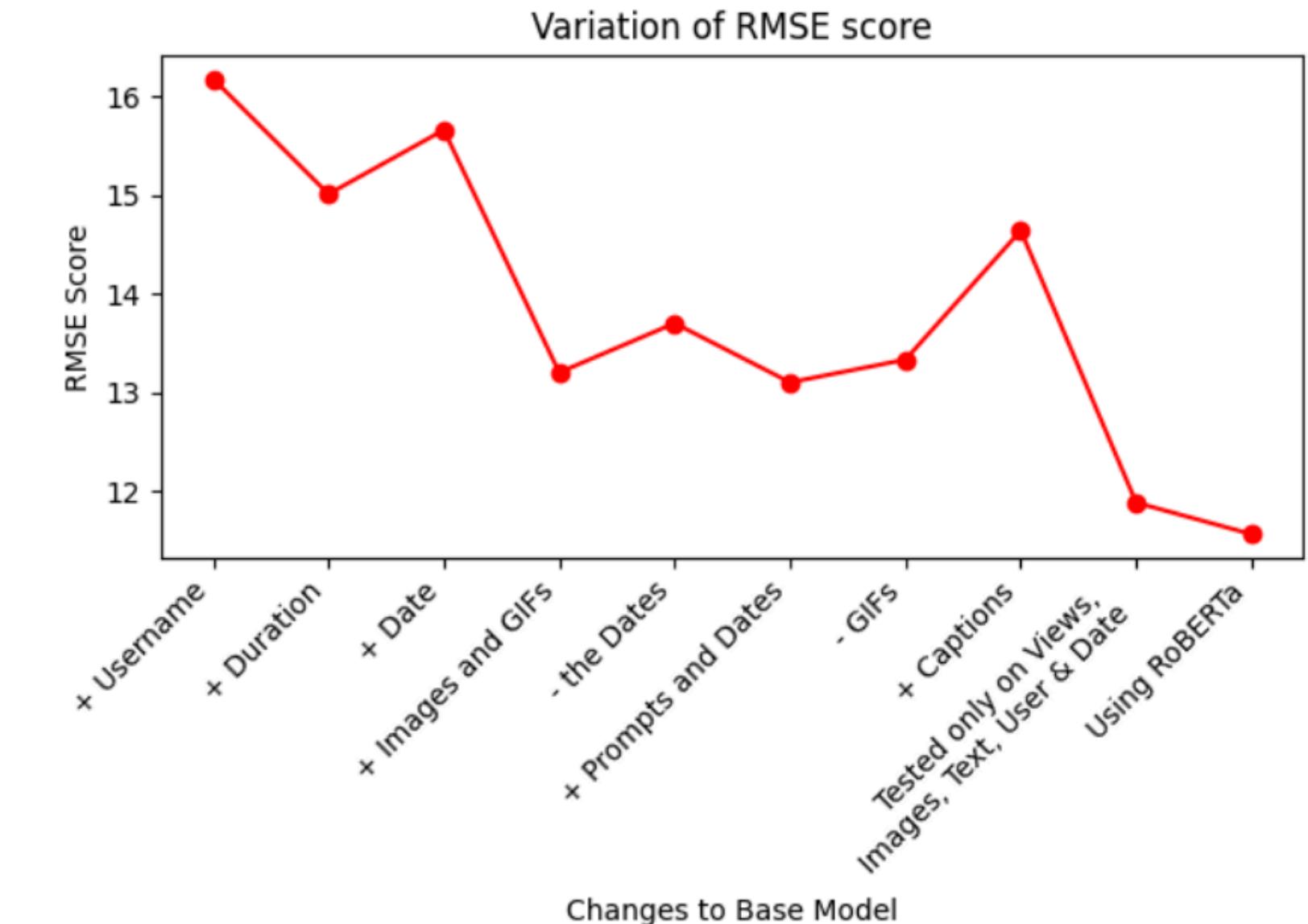
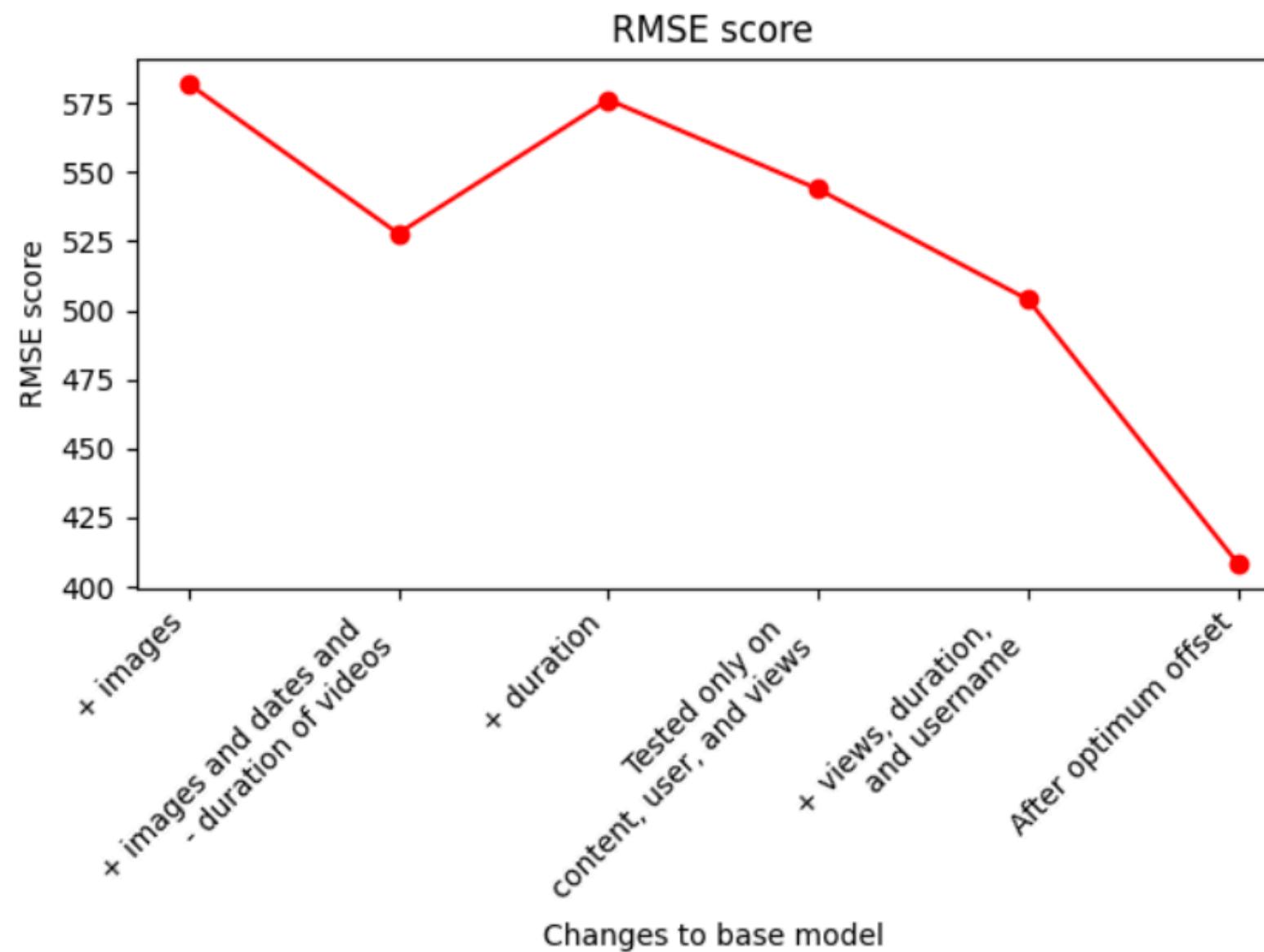
Pronounced left-skew

# Binning & Classification:

## Hierarchical Binary Classifier

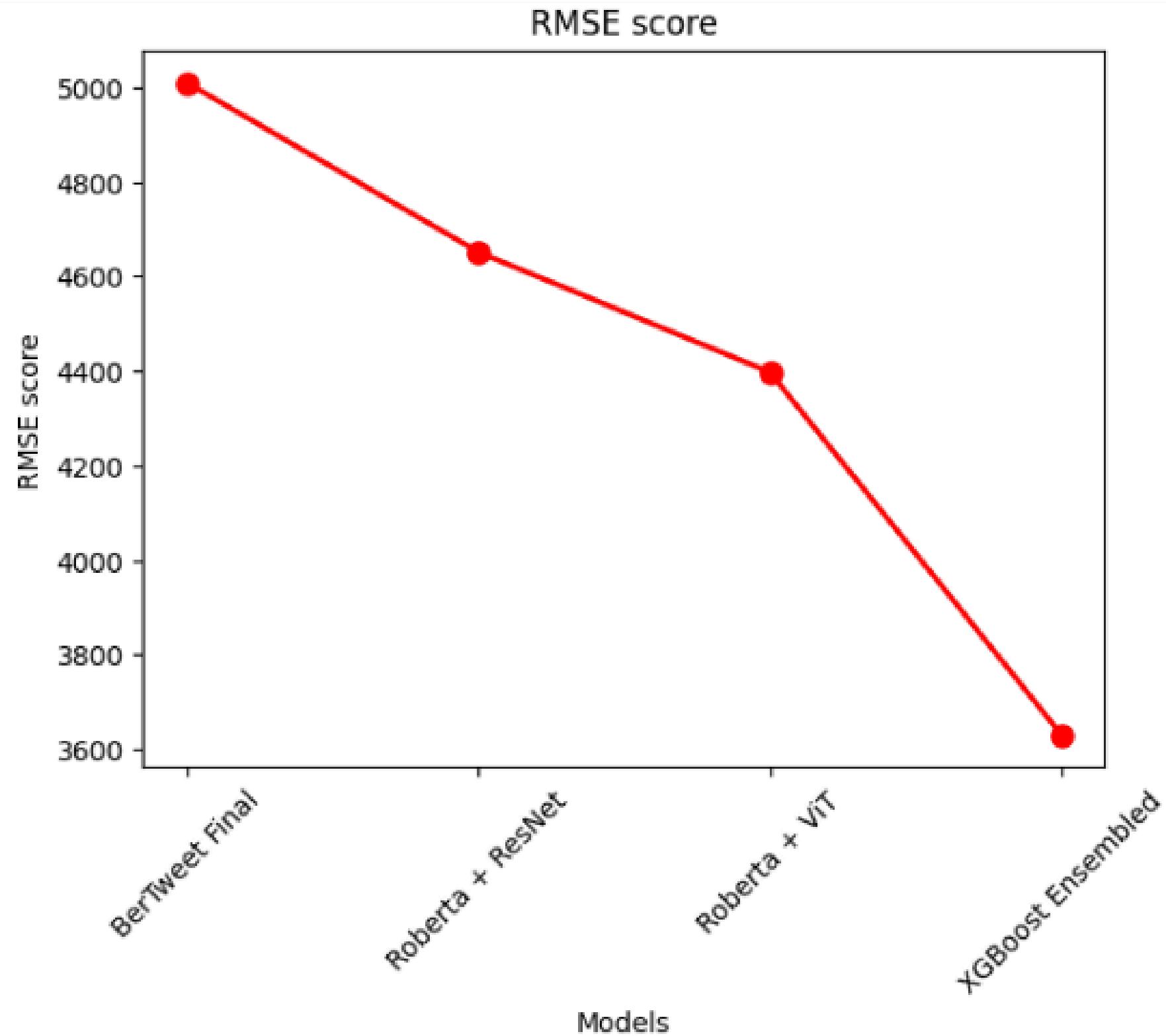


# BERTweet



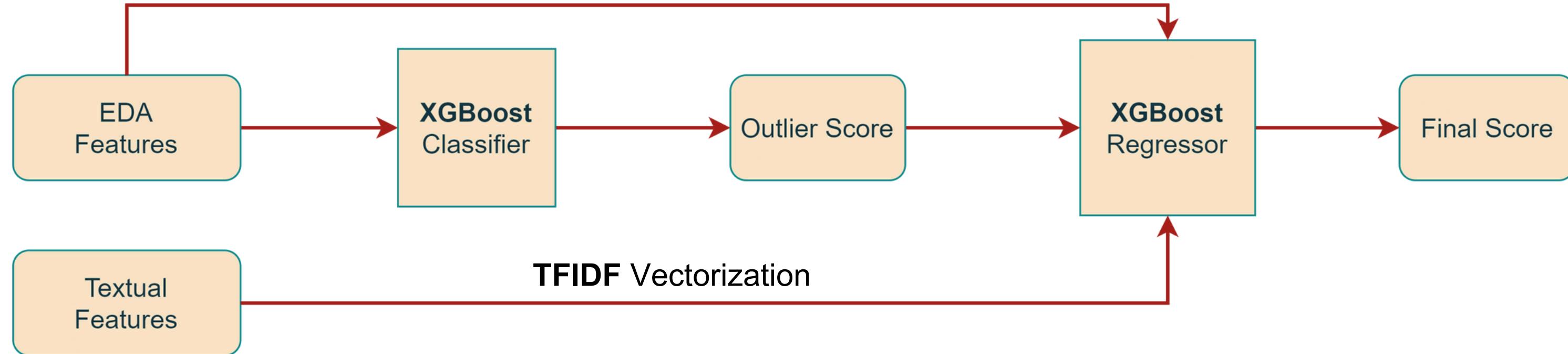
Model Name	Range	RMSE
Bertweet	0-100	11.8
Bertweet not pre-processed	100-3500	508
Bertweet pre-processed	100-3500	408
Bertweet Final	0-∞	5009

# XGBoost



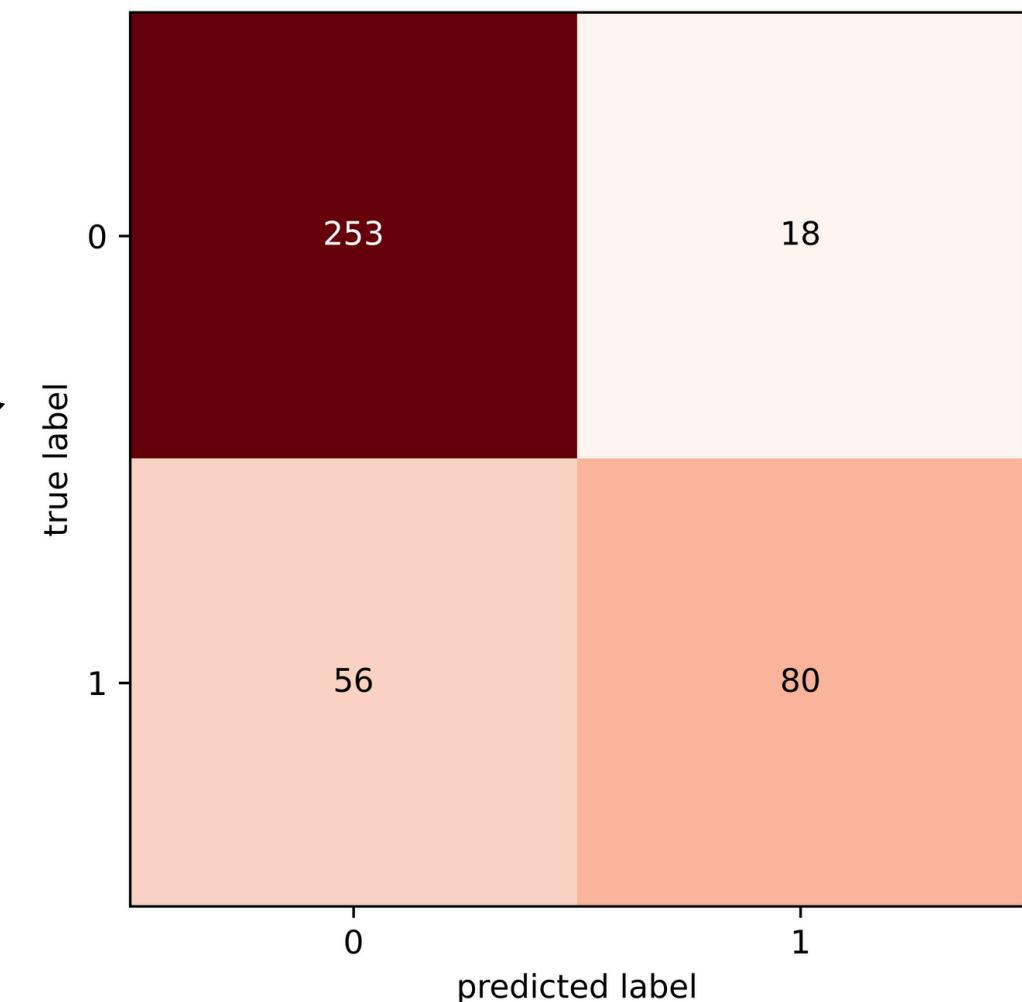
Model Name	RMSE	R2
BerTweet Final	5009	0.02
Roberta + ResNet	4651	-0.1
Roberta + ViT	4397	0.25
<b>XGBoost Ensembled</b>	<b>3631</b>	<b>0.15</b>

# Training Pipeline



Number of parameters are minimal without compromising **accuracy**

The classifier performs with **81% accuracy**



## Approaches and Results

### Early Approaches:

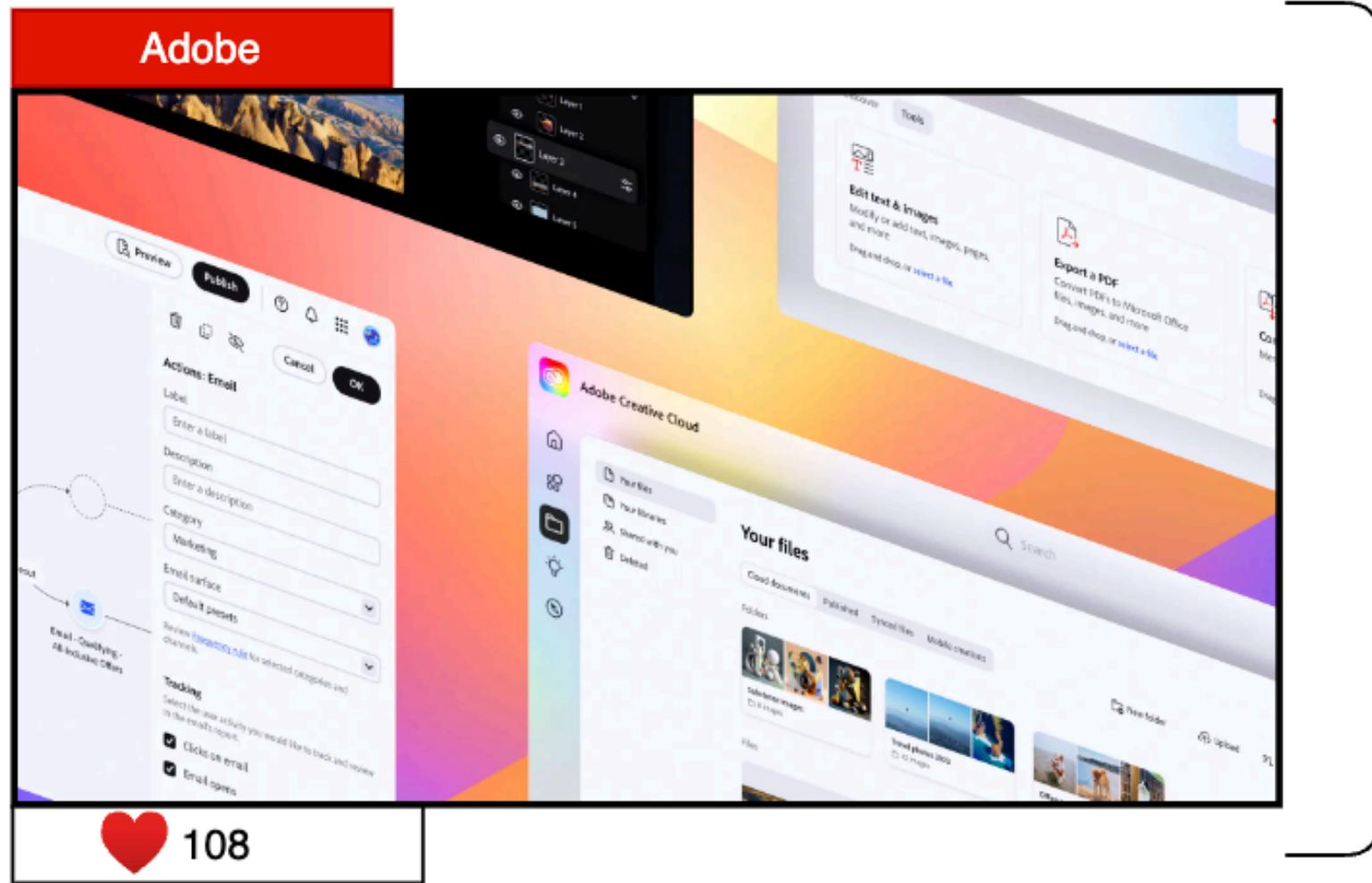
- BERT
- ViT
- Combined-BERT
- ResNet
- BERT Fusion

### BERTweet approaches

Model Name	Range	RMSE	R2 Score
Bertweet	0-100	11.8	0.84
Bertweet not pre-processed	100-3500	508	0.67
Bertweet pre-processed	100-3500	408	0.71
Bertweet Final	0-∞	5009	0.02
XG-Boost Ensembled	0-∞	3880	0.15

However, the output metrics obtained from these complex architectures were found to be **comparable** to those achieved by XGBoost.

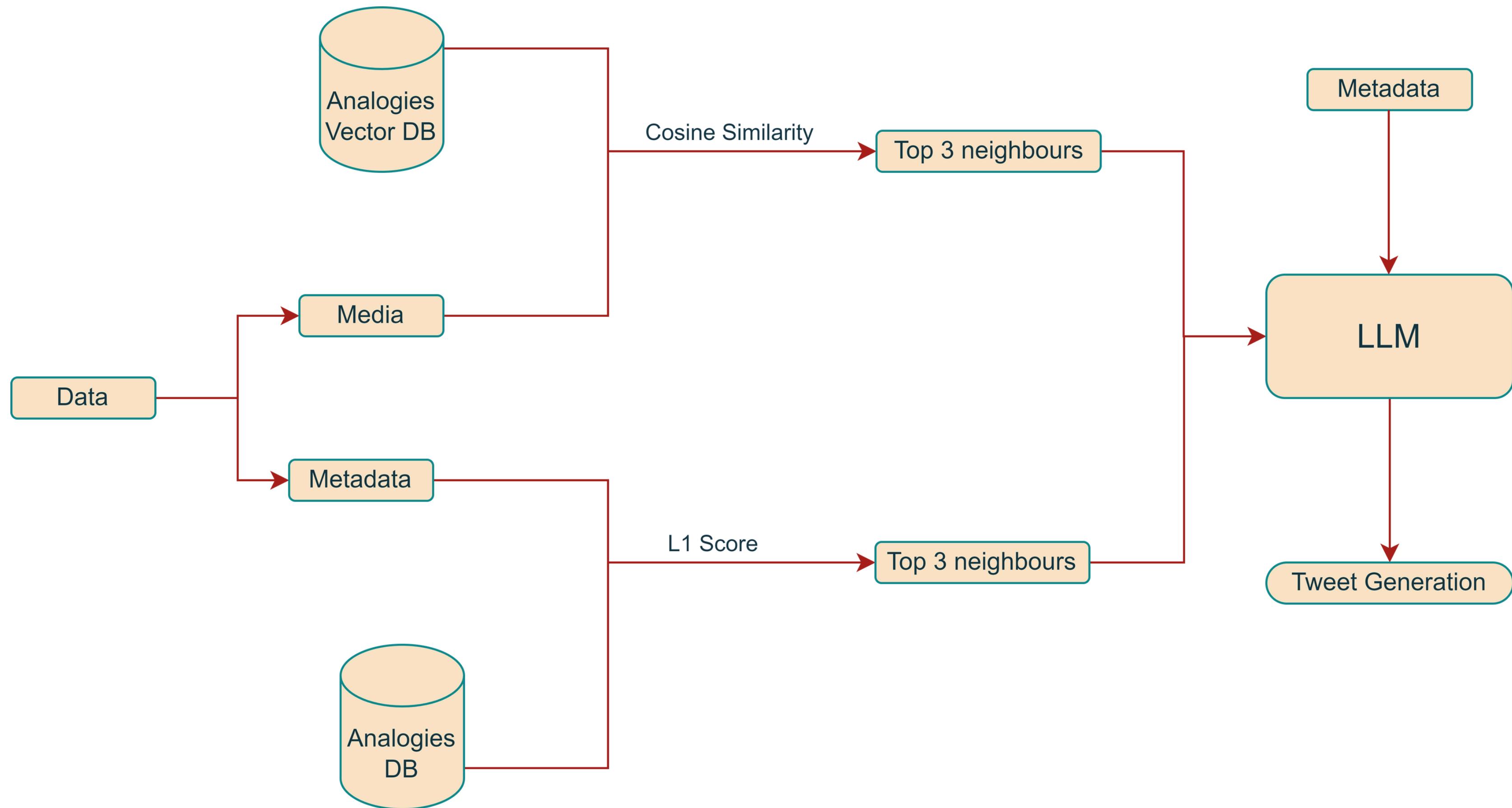
# Task 2: The Challenge



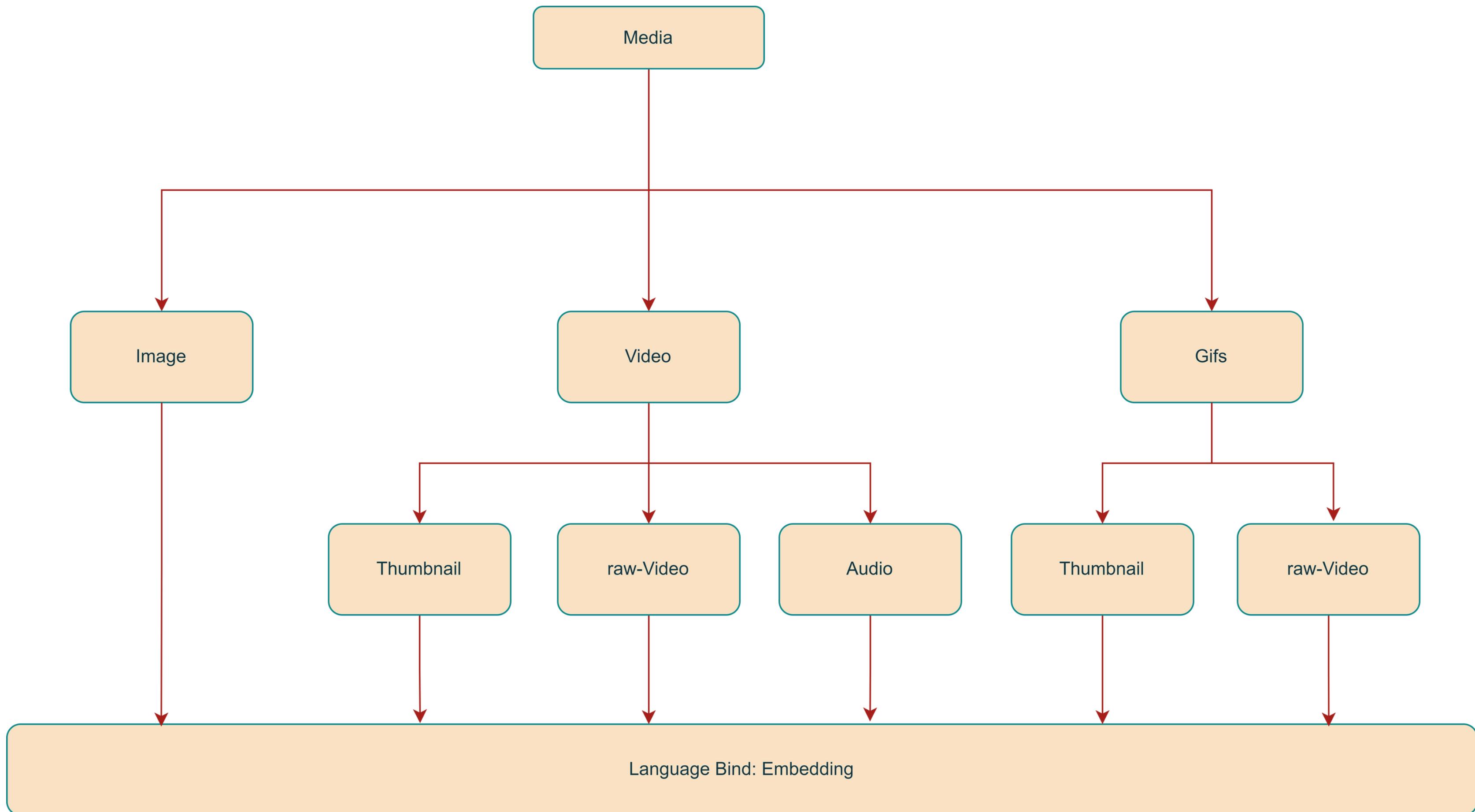
Predict

Say hello to Spectrum 2, a comprehensive update to the design system shaping all of Adobe's tools and experiences.

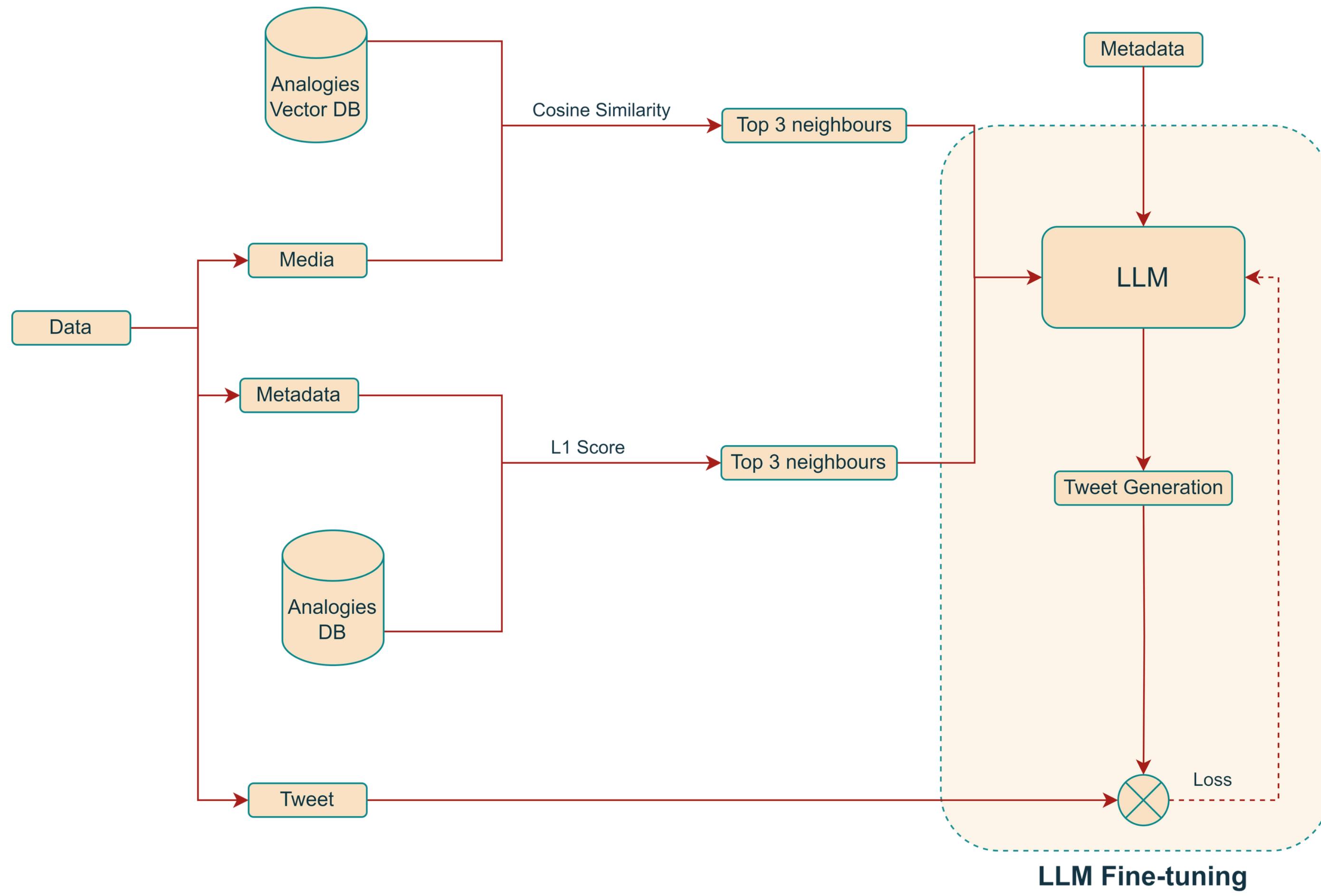
## Task 2: Tweet Generation Pipeline



# Task 2: Encoding Media



# LLM Fine-tuning



## **TASK-2: RESULTS**

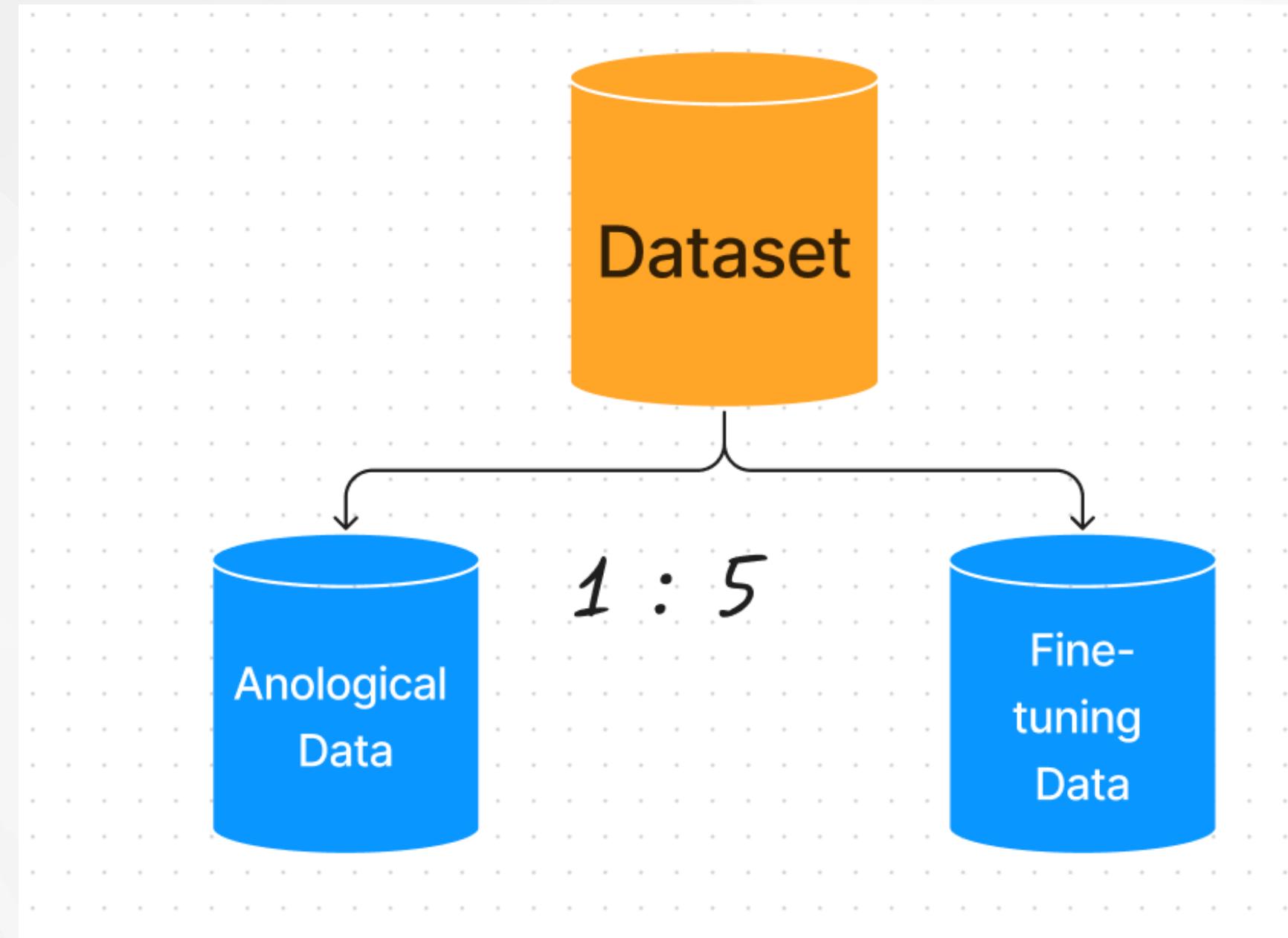
<b>Final Results</b>					
<b>LLM</b>	MISTRAL 7B INSTRUCT - mistralai/Mistral-7B-Instruct-v0.1				
<b>Fine-tuned on</b>	20k datapoints using media and metadata like company name, likes, date and tweet				
<b>Performance Metrics</b>					
<b>bleu_score 1</b>	<b>bleu_score 2</b>	<b>bleu_score 3</b>	<b>bleu_score 4</b>	<b>rouge_score</b>	<b>cider_score</b>
0.11642936	0.04942742	0.02377453	0.01508564	0.11710835	0.07782196

**DEMO**

## **TASK-2**

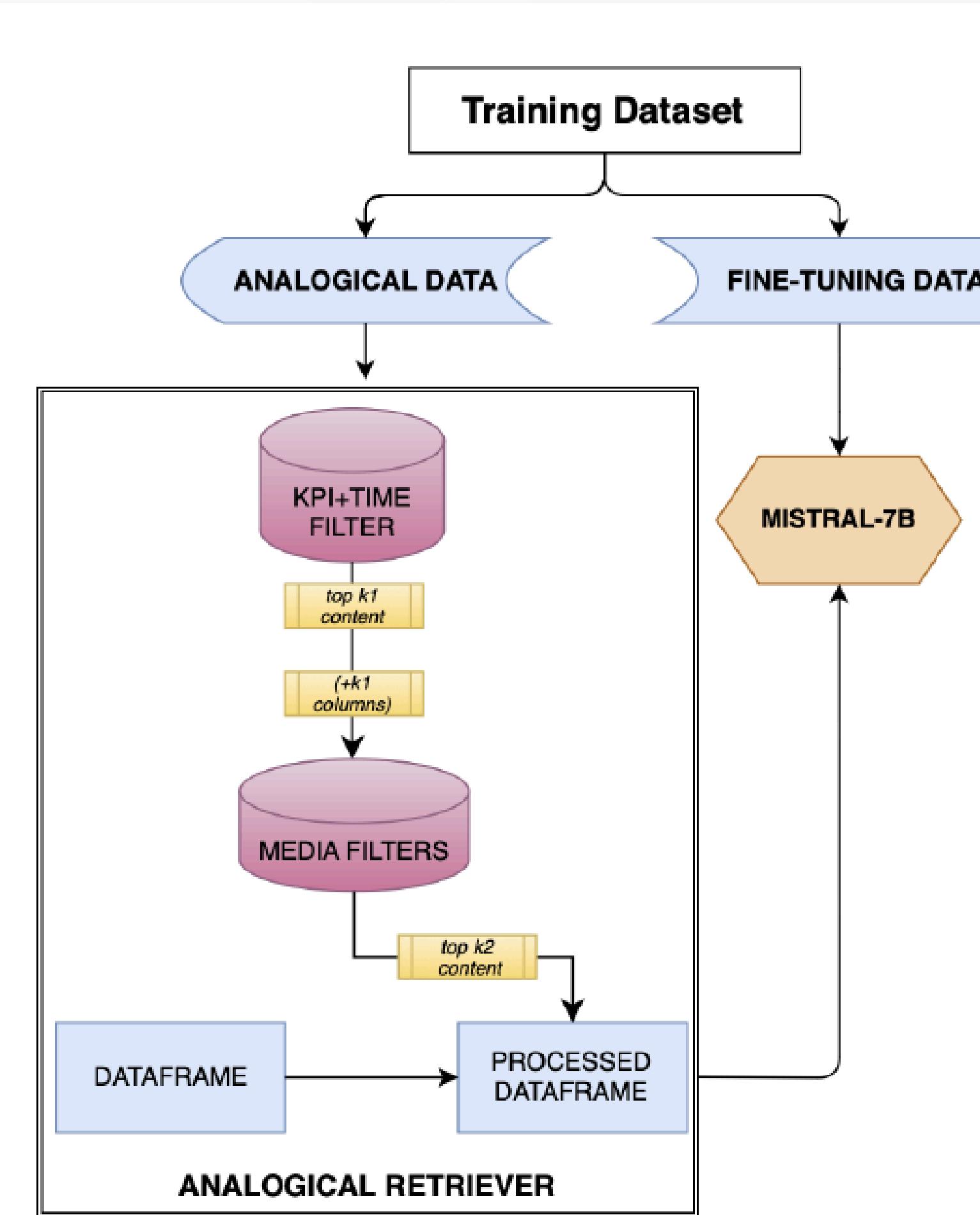
**THANK YOU**

# TRAINING PIPELINE



We add  $k_2$  additional data columns to each row of the analogical data by finding similar data points.

# TRAINING PIPELINE



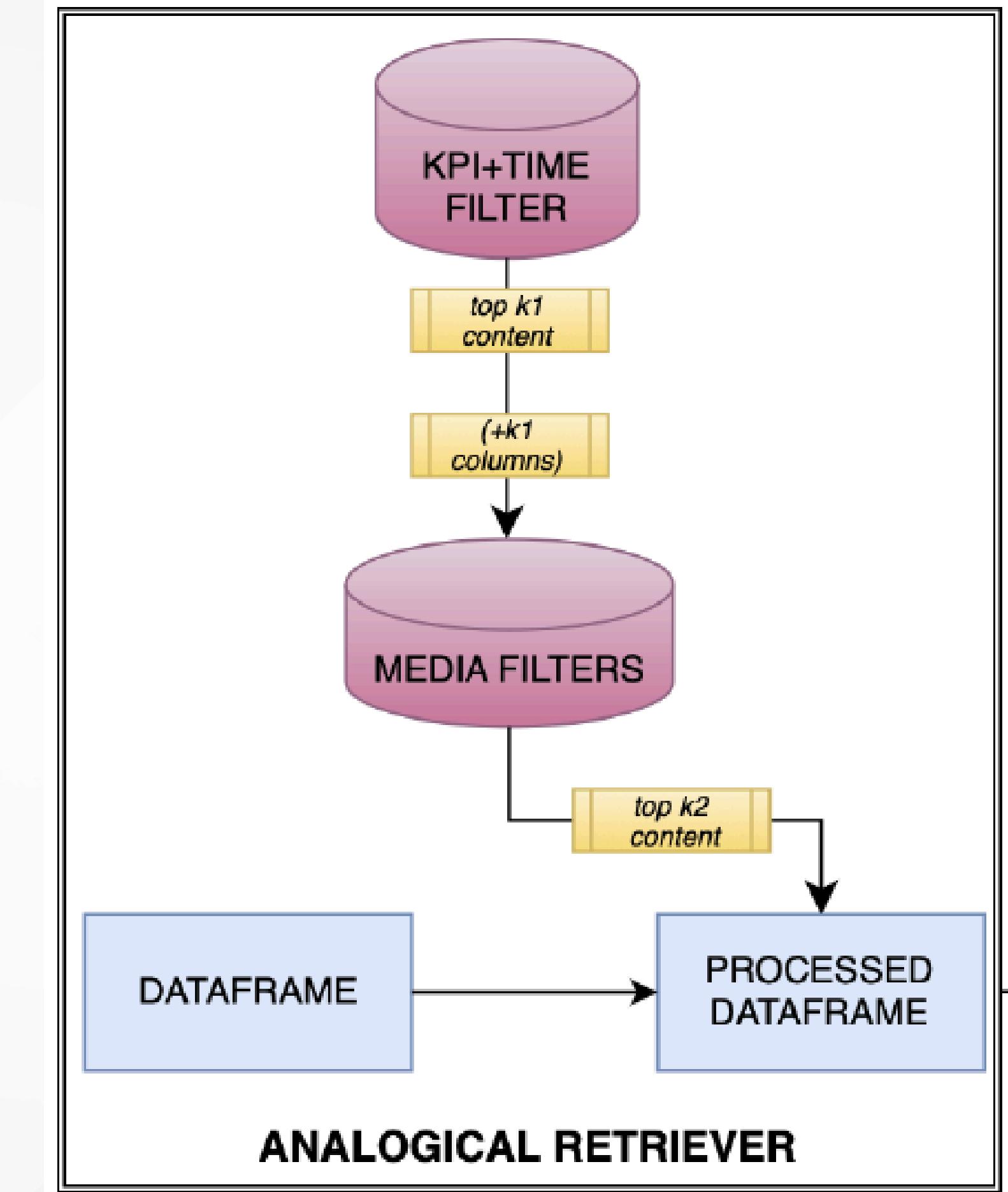
# TRAINING PIPELINE

## Analogical Retrievers :

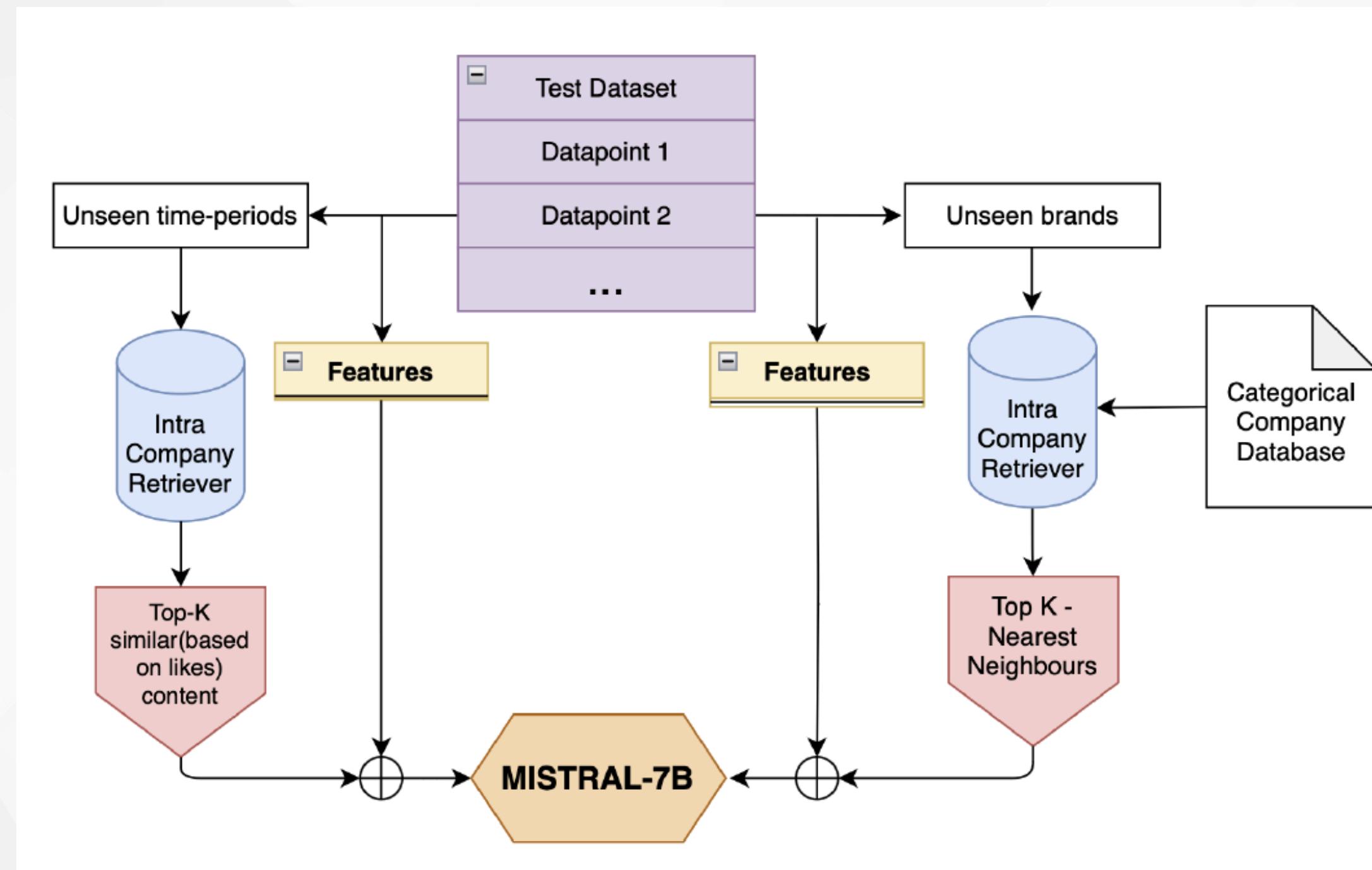
Tweets are sorted by date and likes (**KPI+TIME FILTER**), with the **top k1** tweets selected.

These are further sorted based on **similarity scores** from the generated embeddings.

The final selection includes the **top k2** tweets, which are then added to each datapoint.



# INFERENCE PIPELINE

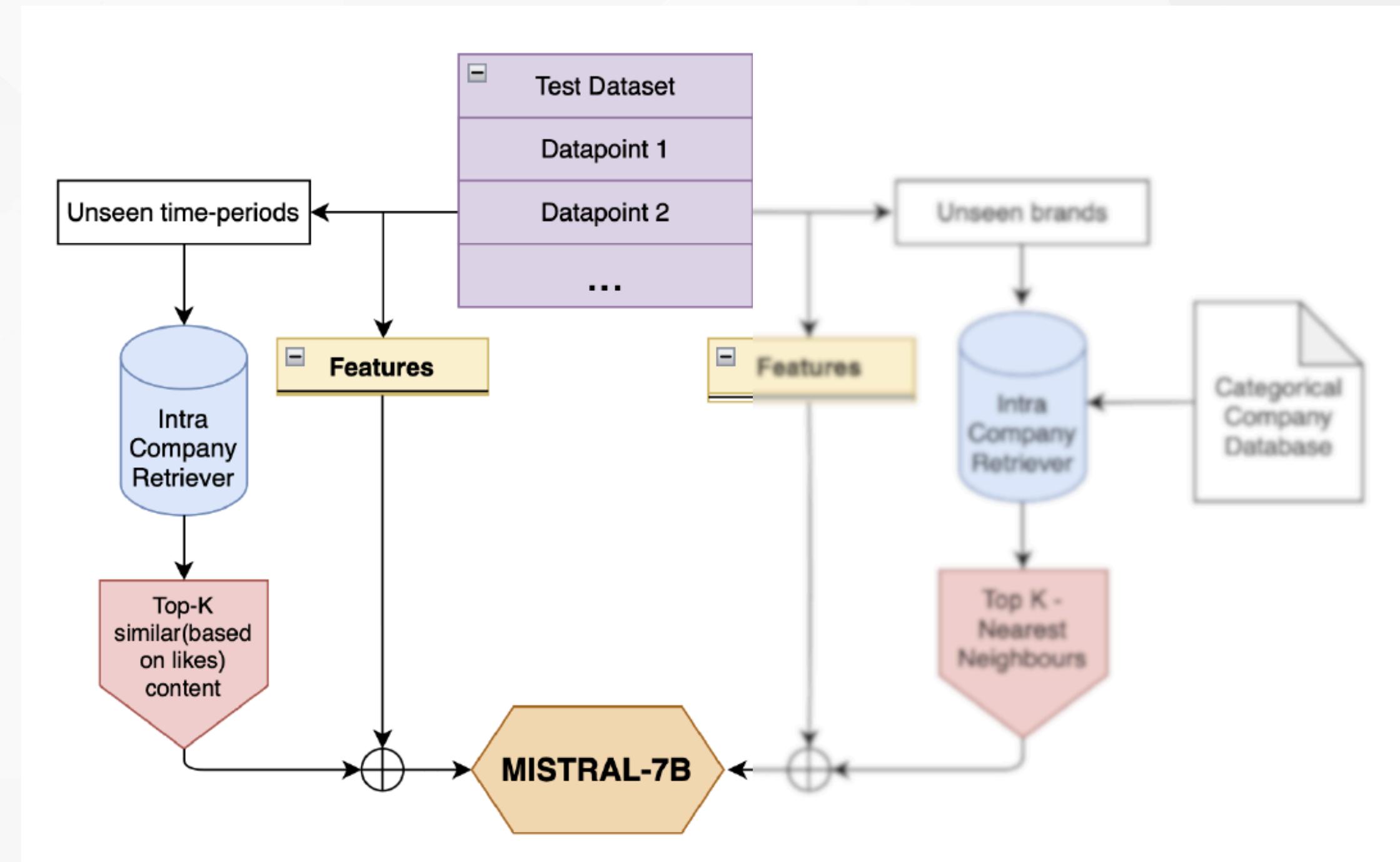


**CUSTOM MISTRAL-7B**

# INFERENCE PIPELINE

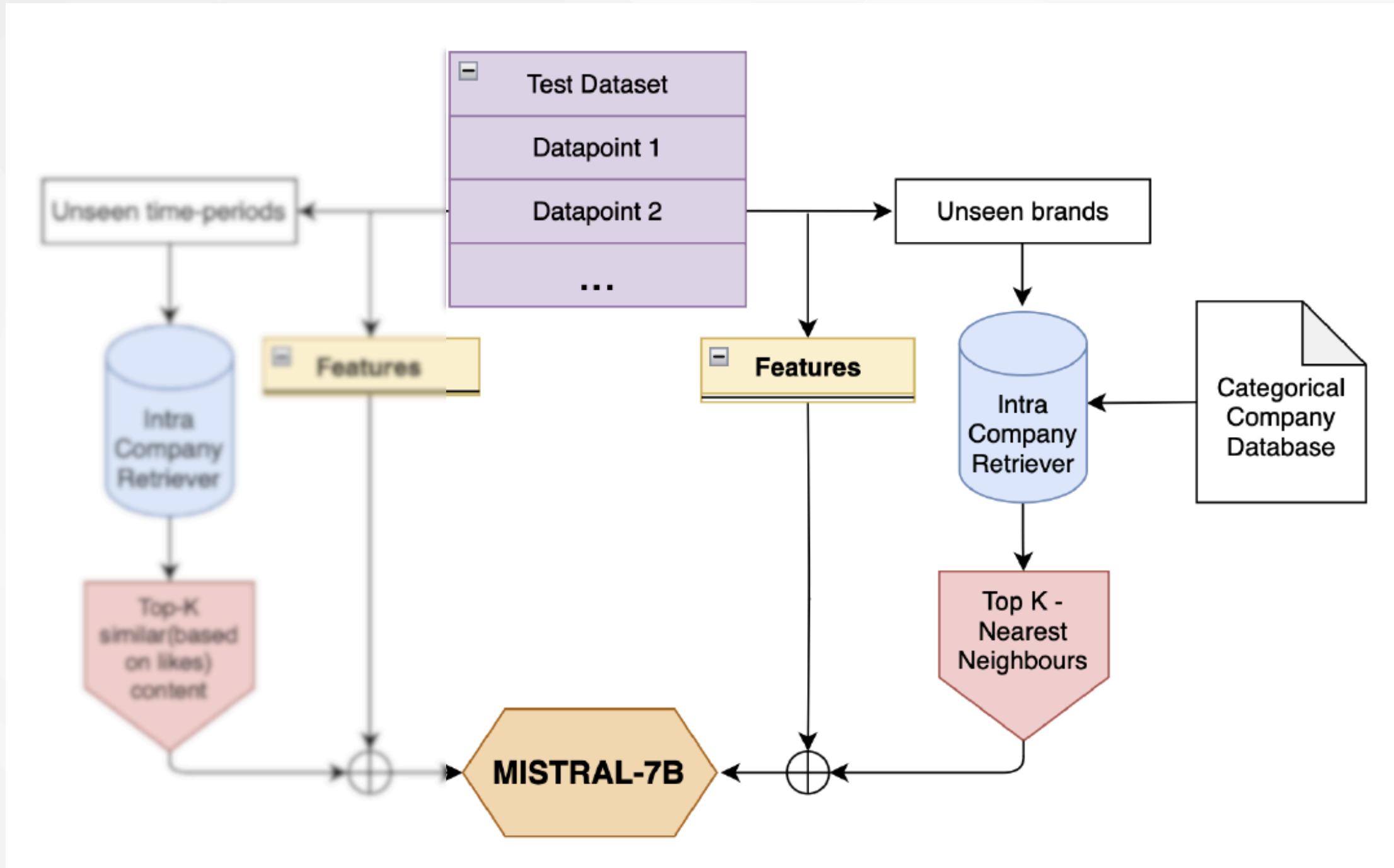
## Unseen Time periods :

Enhancing Data with  
Top K Liked Tweets:  
Adding high-  
engagement tweets  
from associated  
company



## CUSTOM MISTRAL-7B

# INFERENCE PIPELINE



## Unseen Brands:

Gathering top K liked tweets from similar companies

## CUSTOM MISTRAL-7B

# The following prompt was used in the fine-tuning of Mistral 7B

Listing 1: Prompt for Mistral-7B

<s><s> [INST]

Using the given tweets as reference construct a tweet which has analogical similarity to those , conditioned on the fact that the post is a Moderate Engagement post.

Reference tweets:

Tweet: China built a hospital in just 10 days for coronavirus patients <hyperlink>

Username:Independent

Company:independent

Likes:Moderate Engagement

Tweet: A massive construction effort is under way in Golokhvastovo , 50km southwest of Moscow , to build a hospital for the treatment of COVID-19, the disease caused by the coronavirus. The site is expected to have 500 patient beds and separate dormitories to house a staff of 1,000.

<hyperlink>

Username:RFERL

Company:free

Likes:High Engagement

Tweet: The tunnel boring machine "Prerna" has reached its destination at Esplanade Metro station today after a fascinating journey of 3.8 kms below the twin cities of Howrah and Kolkata including the river Hooghly <hyperlink>

Username:metrorailwaykol

Company: metro

Likes:Low Engagement

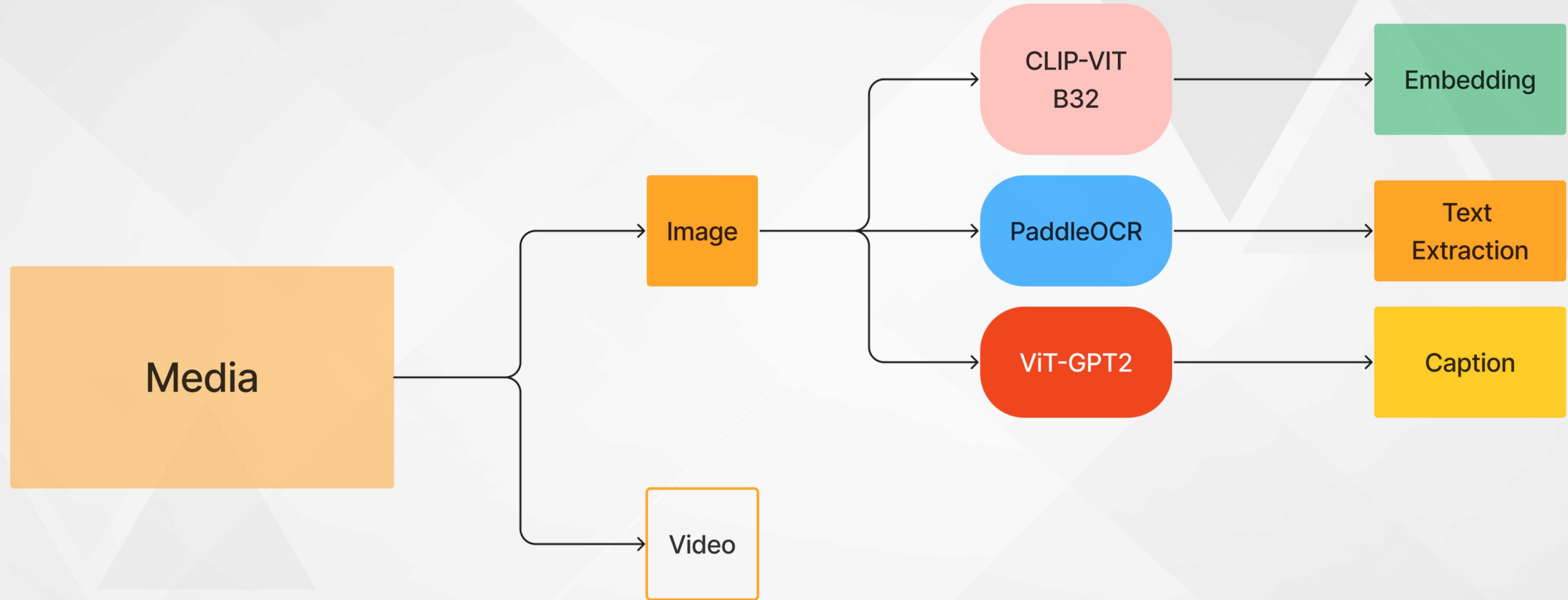
The tweet is written by the user cnni belong to cnn.

New Tweet: An 85-year-old primary school in Shanghai has been lifted off the ground in its entirety and relocated using new technology dubbed the "walking machine."

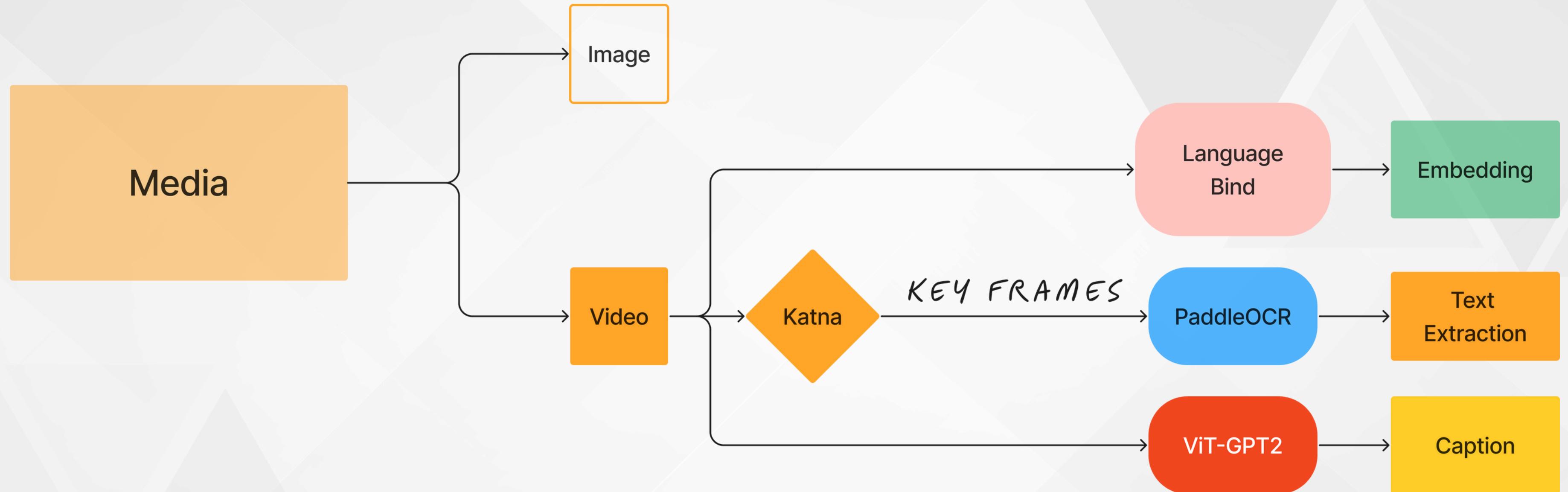
<hyperlink> <hyperlink> </s></s><s>

# **Appendix**

# Image Processing



# Video Processing



# Audio Processing

