A Corpus for Dimensional Sentiment Classification on YouTube Streaming Service

Ching-Wen Hsu, Chun-Lin Chou, Hsuan Liu and Jheng-Long Wu

Department of Big Data Management, Soochow University, Taipei, Taiwan

CONTENTS

Introduction

Habitual

People viewing habits have shifted from TV to online social media platforms.

Discussable

YouTube provides a discussion function for audiences to express their opinion.

Why we use YouTube?

Shareability

Videos themselves are highly shareable, and thus become a fast and effective way to convey ideas.

Omnipresent

People can watch videos at anytime and anywhere.
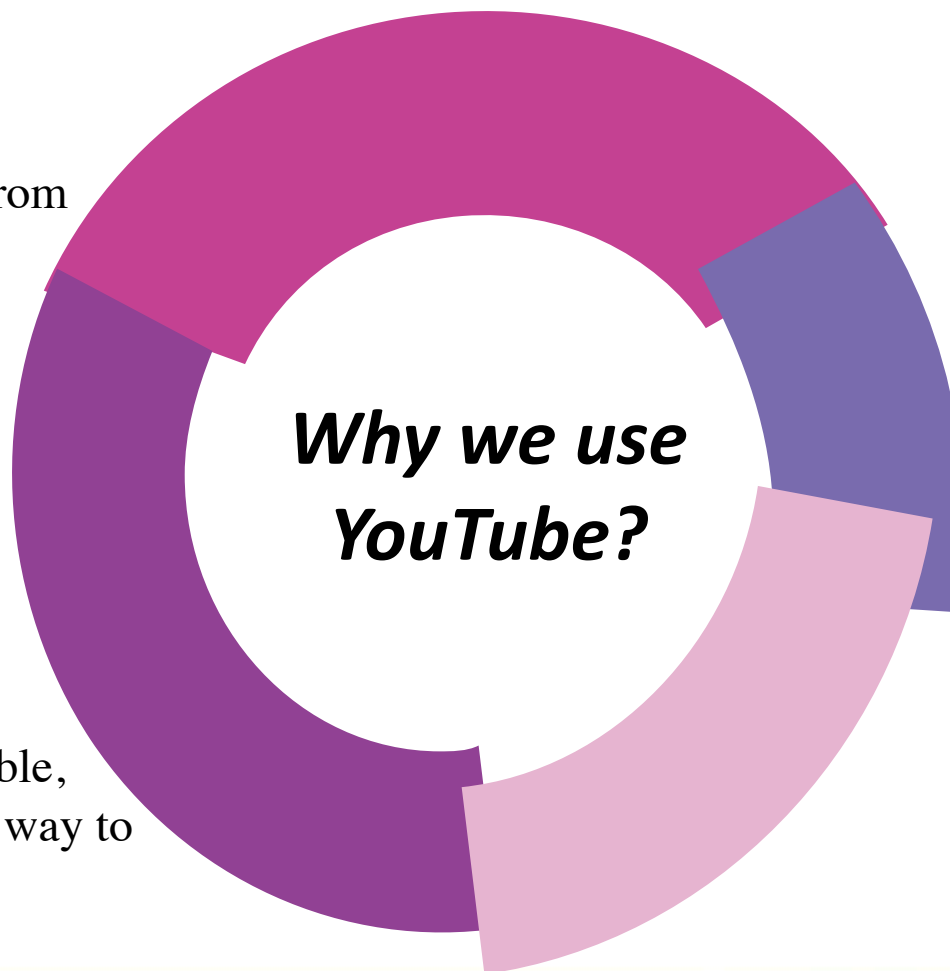
Motivation

How do we analyze video performance ?

Present ways to tract video performance

1. average view duration.
2. browsing history.
3. variance in audience's demographics.

Comments themselves are useful

Analyze audience's comments through multi-dimensional sentiment indicators.

Help YouTuber create improved content

1. give YouTubers a sense of its popularity.
2. see what resonates with their audience.
3. discover audience emotional ups and downs.

Related Work

The purpose of analyzing YouTube's comments

Clustering comments to classify video categories (Leung et al., 2009).

01

Social media reflect public views

Using Amazon's comments to determine audience's preference (Bhatt et al., 2015).

02   **03**

methods deal with text-based classification tasks

Fine-tuned with Bidirectional Encoder Representations from the Transformers (BERT) model using comments.
(Sun et al. 2019)

Methodology

Comment Collection

Sentiment Indicators

Data Labeling

Text Preprocessing

Text Classification

25 channels

125 videos

12,500 comments

Three indicators for

analyzing comments

1. YouTuber preference

2. Video preferences

3. Excitement level

Three experts annotate
sentiment indicators .
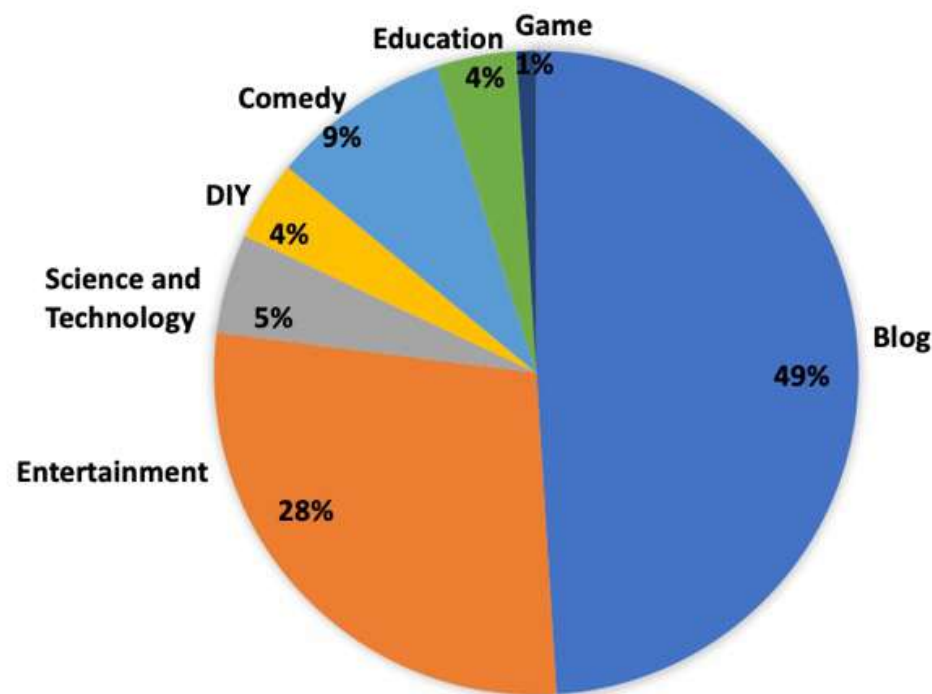
TF-IDF
word-embedding

RandomForest
Xgboost
SVM
FastText
BERT

Comment Collection

Channels

- Subscribe number over 100,000
- Highly popular or controversial.
- A total of 25 channels.
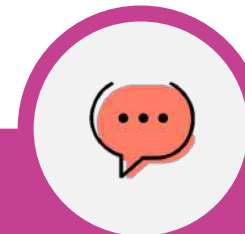
Comment Collection

Channels
- Subscribe number over $100,000$
- Highly popular or controversial.
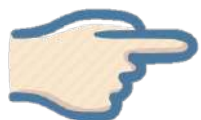- A total of 25 channels.

videos
- selection video duration 2019/01/01-2021/02/31
- filter top 5 videos through viewing number.
- A total of 125 videos.

Comments
- randomly remain 100 pieces of comments from each videos
- A total of 12,500 comments

Sentiment Indicators

YouTuber Preference

Non-Relative Comments towards YouTubers

comments do not contain YouTuber's name or affairs.

STEP 1

Comments Collection

Relative Comments towards YouTubers

STEP 2

Like

Neutral

Unlike

Sentiment Indicators

Video Preference

Non-Relative Comments towards Videos

comments do not contain video content or talk about YouTuber's affairs.

STEP 1

Comments Collection
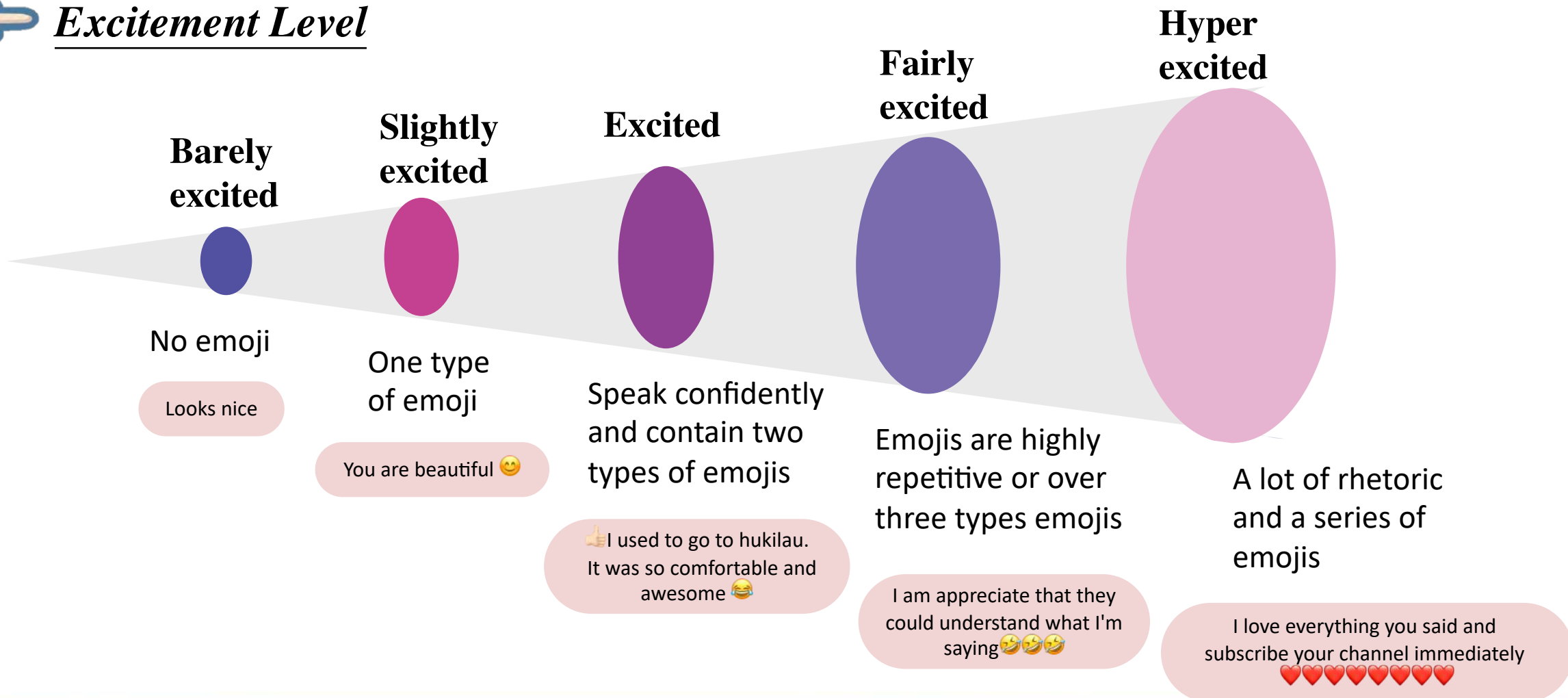
Relative Comments towards Videos

STEP 2

Like

Neutral

Unlike

Sentiment Indicators

Excitement Level

Hyper excited

Fairly excited

Excited

Slightly excited

Barely excited

No emoji

Looks nice

One type of emoji

You are beautiful 😊

Speak confidently and contain two types of emojis

👍I used to go to hukilau. It was so comfortable and awesome 😂

Emojis are highly repetitive or over three types emojis

I am appreciate that they could understand what I'm saying🤣🤣🤣

A lot of rhetoric and a series of emojis

I love everything you said and subscribe your channel immediately ❤️❤️❤️❤️❤️❤️❤️❤️

Data Labeling

watching videos
before labeling
comments is required

label the audience's
comments independently
according to three indicators

watching
videos

annotation guideline

eliminate
non-relative comments

annotation
agreement

Three
annotator

1. advertisements.
2. comments not using Mandarin.
3. comments that post links to external web pages.

using the majority
decision to filter out
inconsistent labels.

Annotation agreement scores

Text Processing

Machine Learning

TF-IDF

Deep Learning

pre-trained word embedding

BERT

pre-trained tokens

Input text

Model training & evaluation

Convert emojis to text

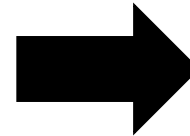Establish a user-defined dictionary

Word tokenization using CKIP
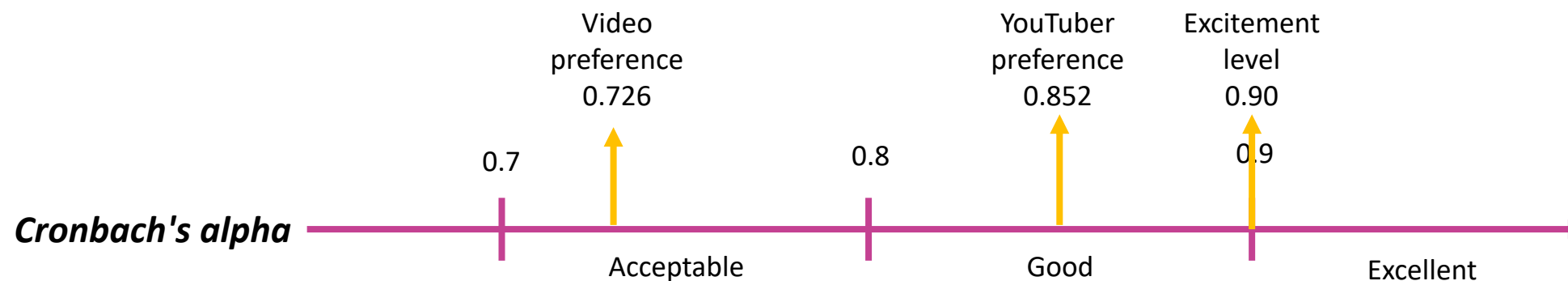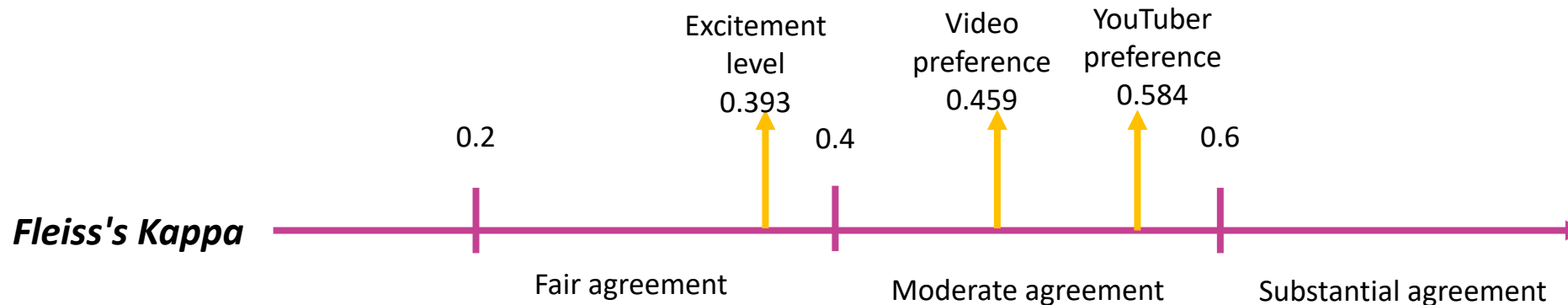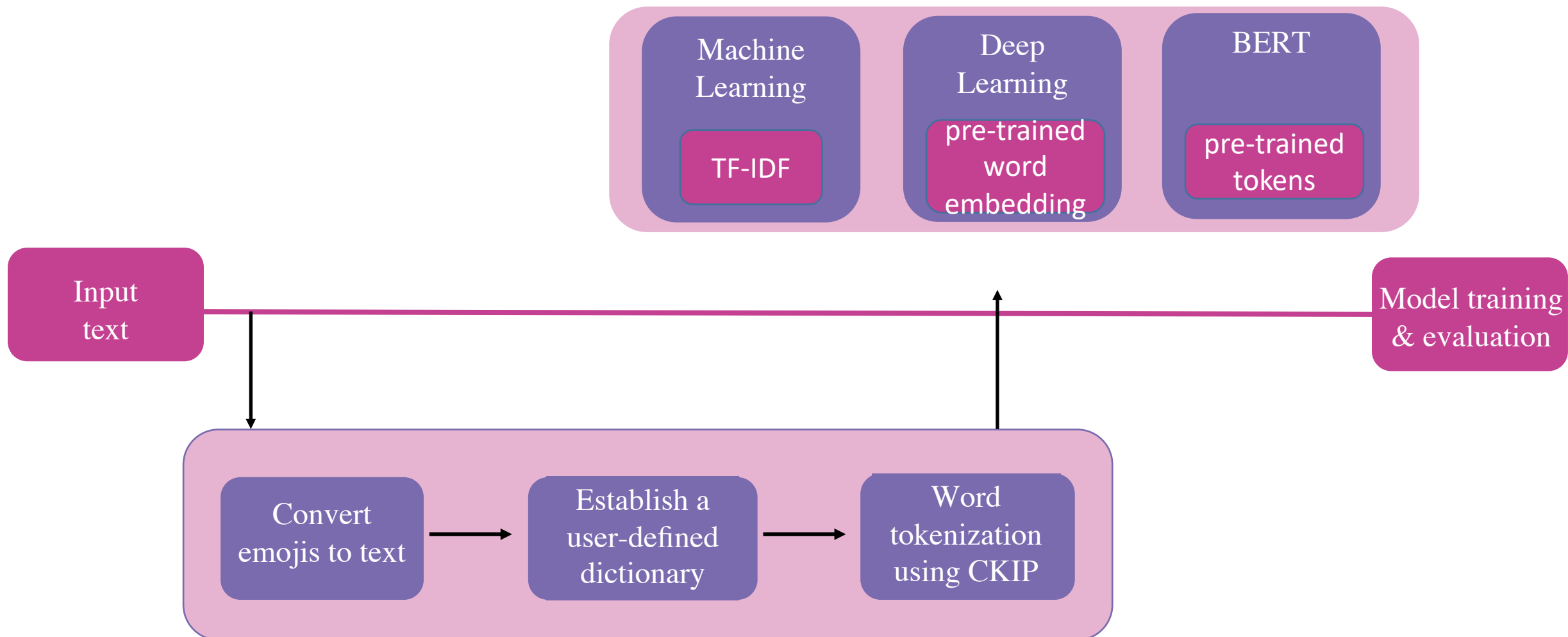
Text Classification

Text Classification

- **BERT-based classifiers**
  - M1: bert-base-multilingual-cased
  - M2: distilbert-base-multilingual-cased
- **Machine learning-based classifiers**
  - M3: RandomForest + TFIDF
  - M4: Xgboost + TFIDF
  - M5: SVM + TFIDF
- **Deep learning-based classifiers**
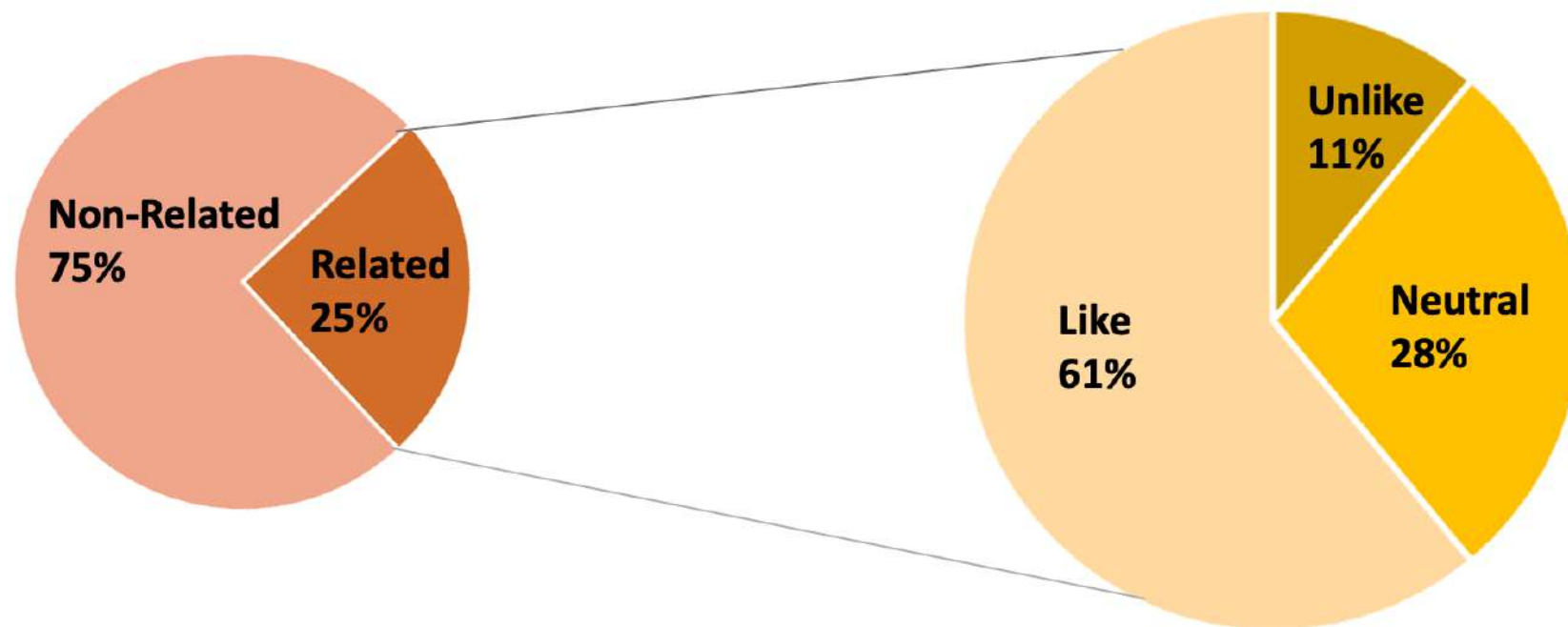 - M6: FastText + Word embedding

Classification Tasks

- T1: Audience's sentiment toward YouTubers.

- T2: Audience's sentiment towards videos.

- T3: Audience's emotional ups and downs.

Experiment

👉 *T1 : Audience's sentiment toward YouTubers*

Data Distribution

Non-Related 75%

Related 25%

Like 61%
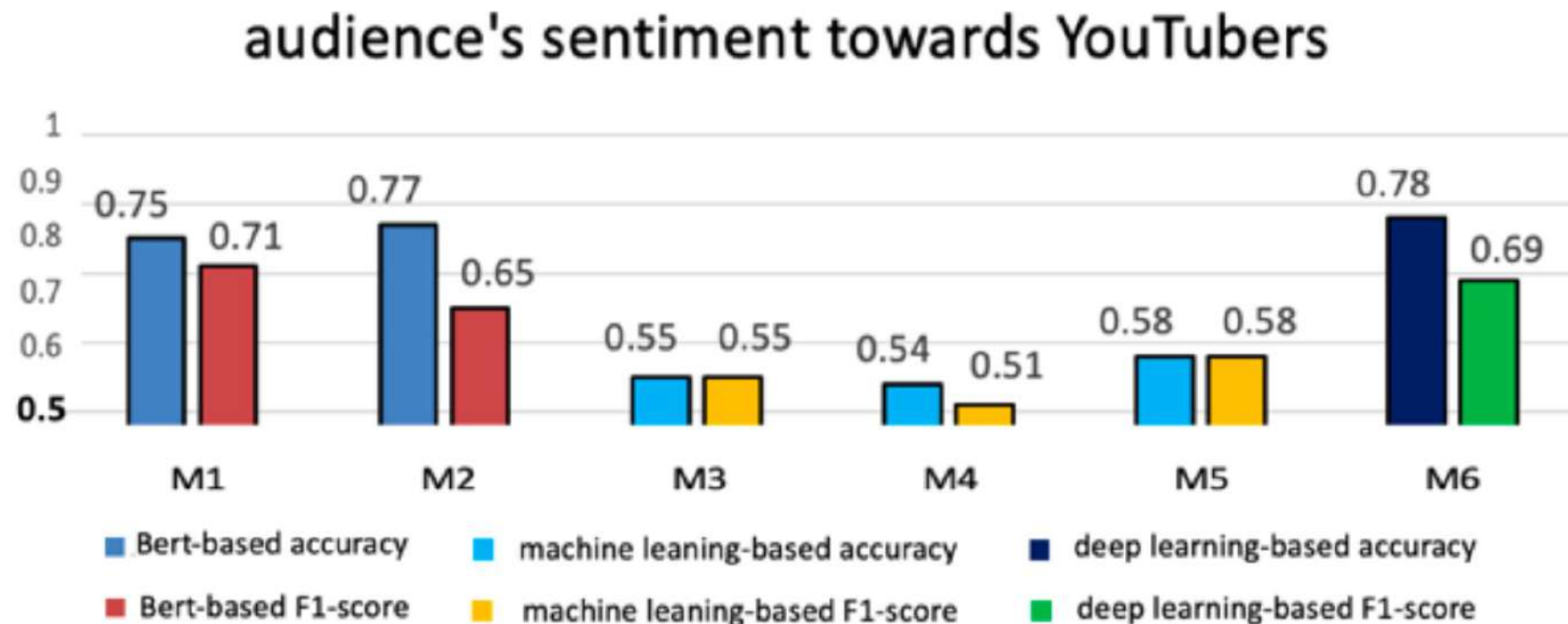
Unlike 11%

Neutral 28%

Over half of the audience present a positive attitude towards YouTubers.

Experiment

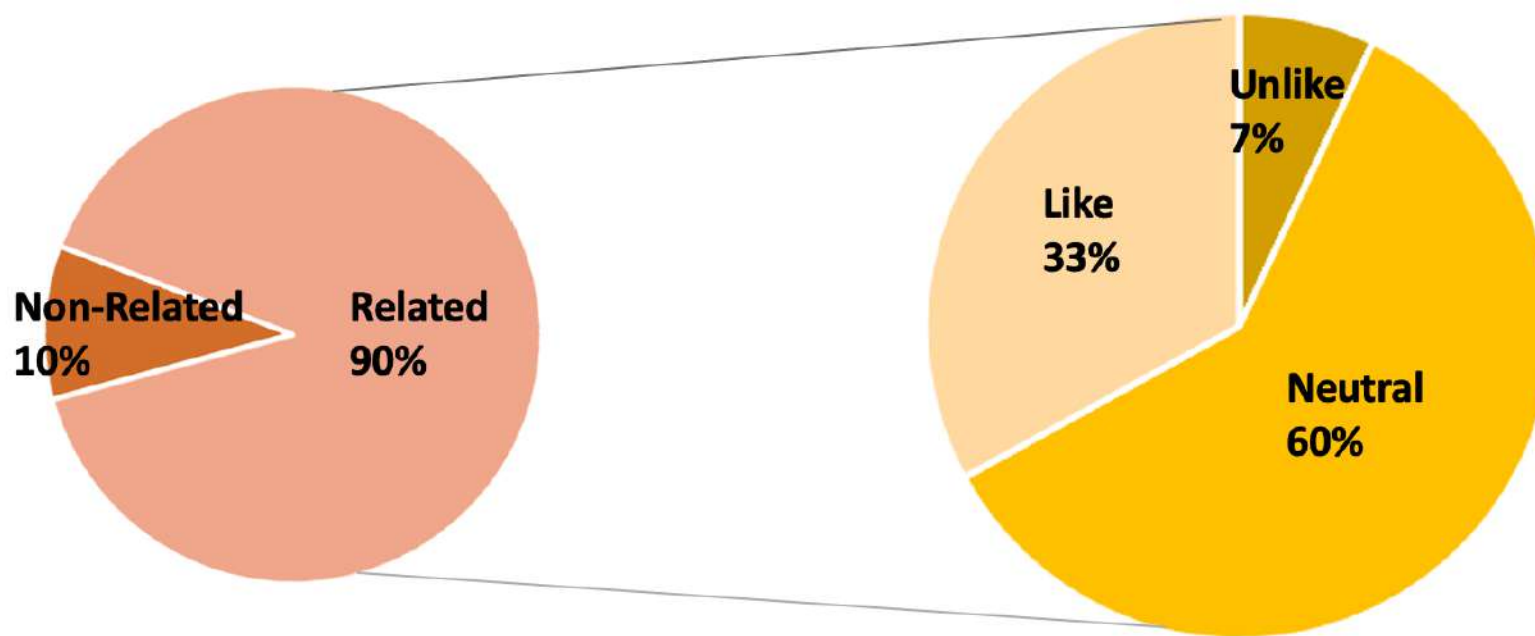👉 *T1 : Audience's sentiment toward YouTubers*

Experiment Result

audience's sentiment towards YouTubers

Bert-based accuracy | machine leaning-based accuracy | deep learning-based accuracy
Bert-based F1-score | machine leaning-based F1-score | deep learning-based F1-score

BERT-based classifiers and deep learning-based classifier have similar performance.

Experiment

👉 *T2 : Audience's sentiment towards videos*

Data Distribution

Although 90% comments have relationship with videos' content, most people do not present their personal positions.
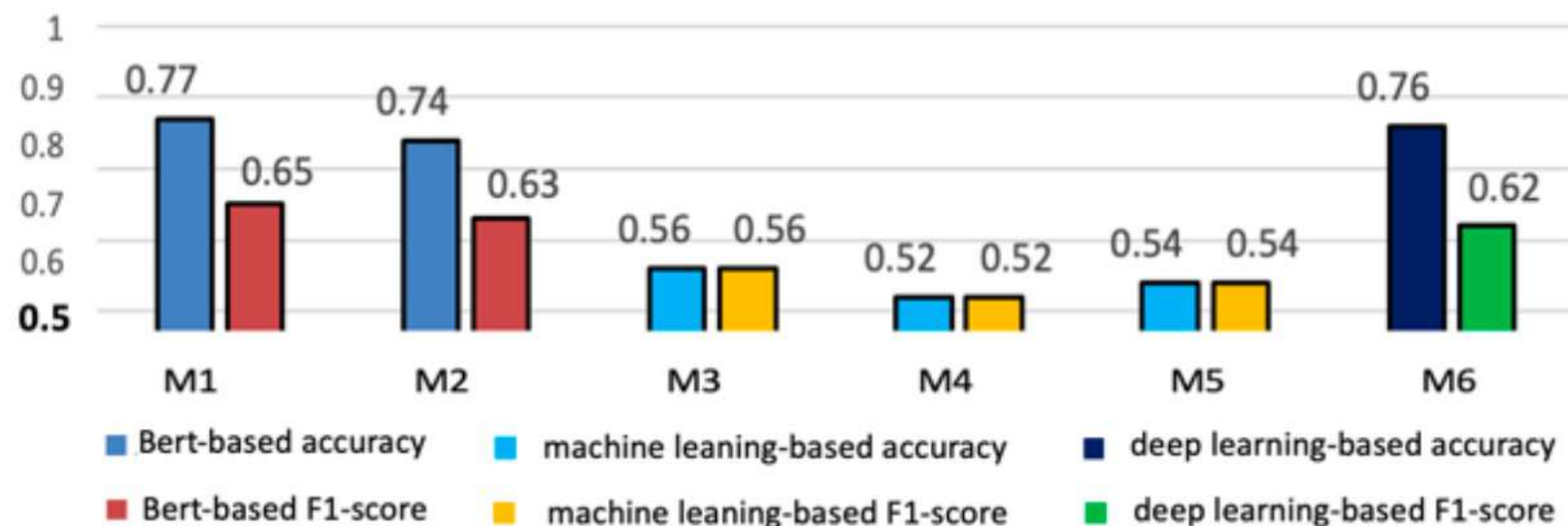
Experiment

👉 *T2 : Audience's sentiment towards videos*

Experiment Result

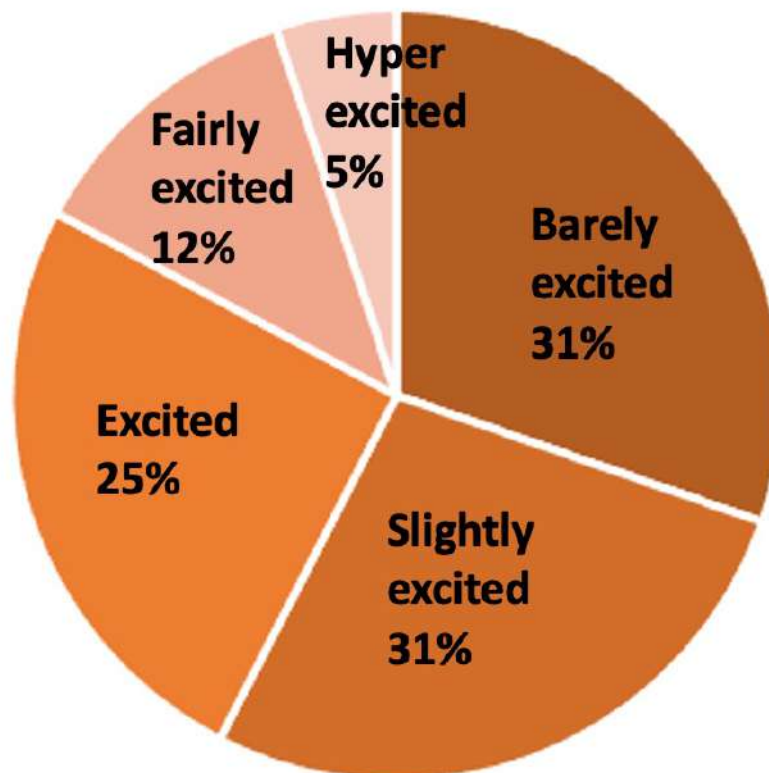audience's sentiment towards video

1. Machine learning-based models achieve the same score in accuracy and F1-score.
2. The accuracy of BERT and deep learning- based methods is 10% higher than their F1-score.

Experiment

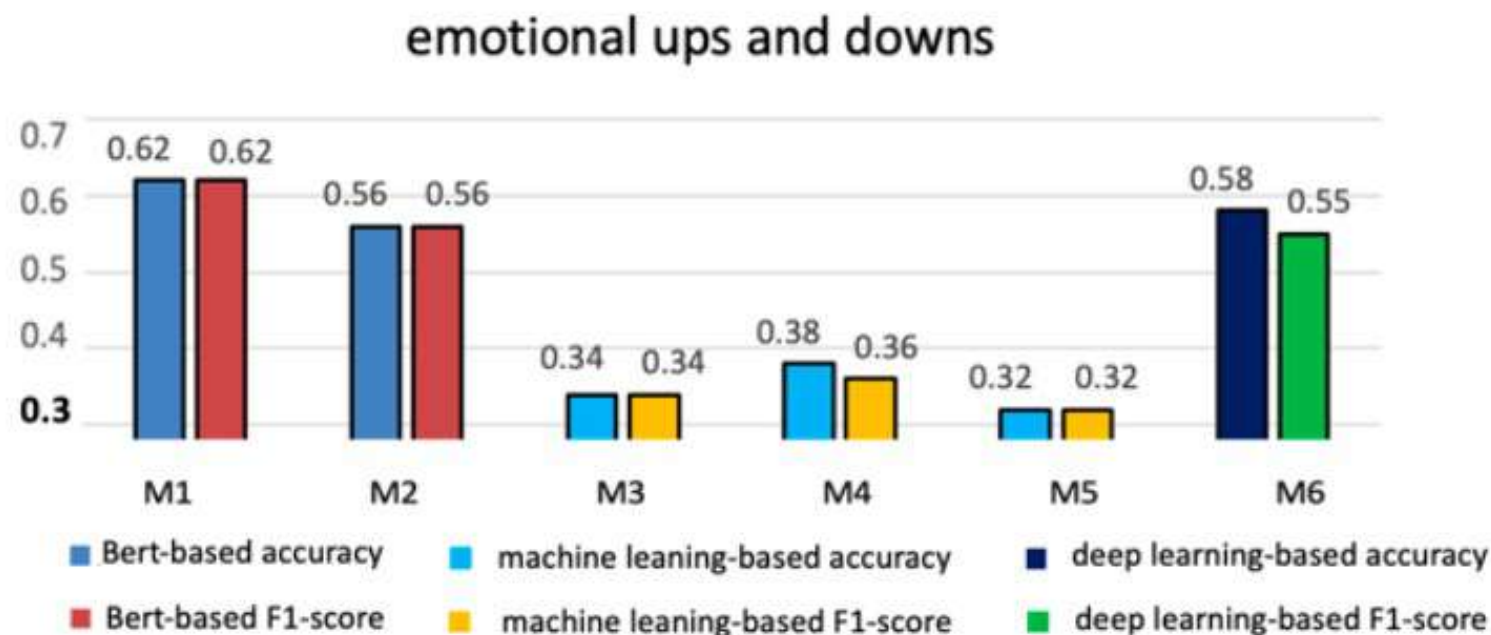*T3 : Audience's emotional ups and downs*

Data Distribution

While Barely excited and Slightly excited account for the majority, some audience also express their extremely emotion.

Experiment

👉 *T3 : Audience's emotional ups and downs*
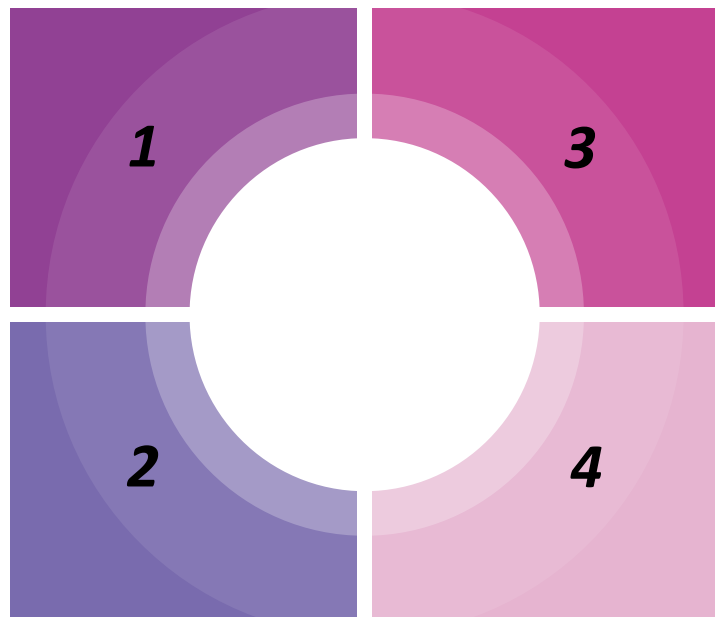
Experiment Result

emotional ups and downs

1. The performance of machine learning-based classifiers reduce significantly.
2. Multi-dimensional tasks are more suitable to be analyzed by BERT-based classifiers and deep learning-based classifier.

Conclusion

Within three sentiment detection tasks, machine learning-based classifiers perform the worst.

1

The majority of comments are related to videos' content; only few of them have relevant with YouTubers.

3

BERT slightly outperforms other models in three tasks according to the F1-score.

2

4

Whatever comments, people not frequently showing negative or even animosity attitude as their opinion.

The end