



SENTIMENT ANALYTICS REPORT

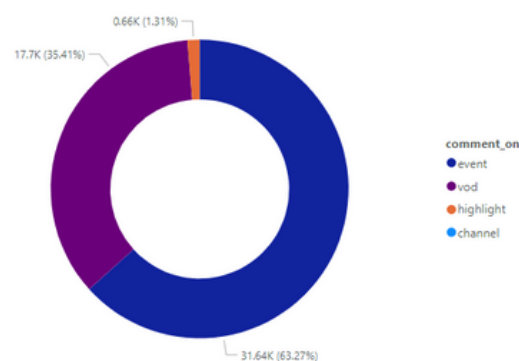
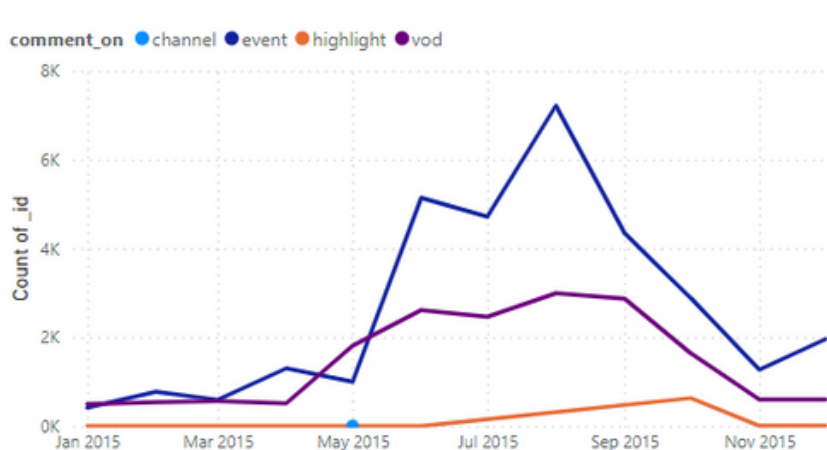
PREPARED BY
TRAN XUAN TIEN

DATA OVERVIEW

Number of comments	Number of users	Avg Comment per user	Avg comments per videos	Number of contents
50K	26.283K	1.90	14.62	3.421K

The data is about the comments of the viewers on our FPT Play platform which are recorded in 2015. Here are some overall views from the data:

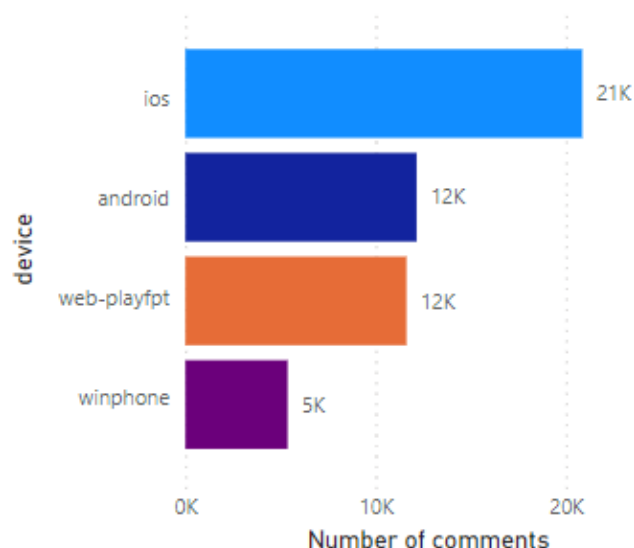
- Total number of comments: **50,000** comments
- Number of users that comments: **26,283** users
- Number of contents (videos/events): **3421** contents
- Average numbers of comments per content: **14.62** comments/content
- Average numbers of comments per users: **1,9** comments/user



Number of comments by Time and Category

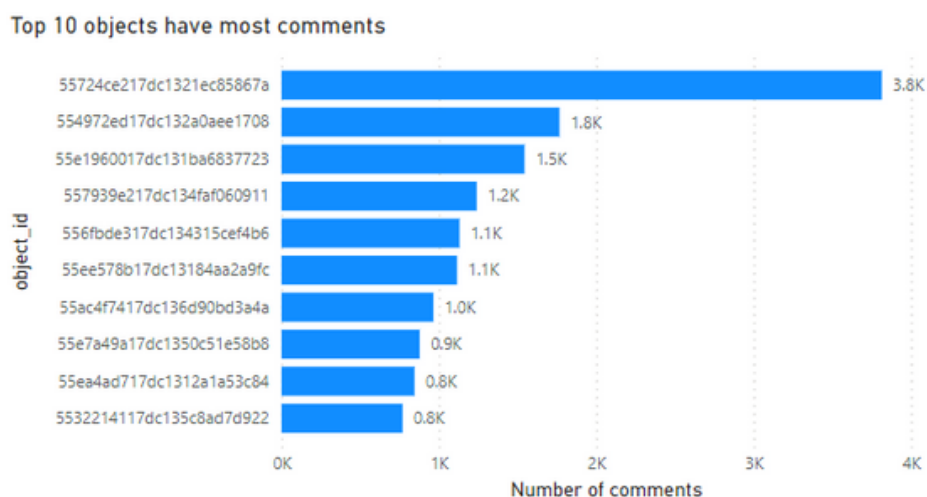
- Users' comments are mostly on the Event (Live Broadcast), and Video on Demands (VOD)
 - There was a rise of number of comments in June 2015 and August 2015, and peak value is nearly 8000 comments for events and about 3000 comments for VODs in August 2015 .
- => **This rise in users' interactions could be caused by the affect of Sea Games 28. Thus, most of the comments of users in on the live events of sport broadcast.**

DATA OVERVIEW



Number of comments by Device Types

- Three main types of device that our users are using to comments are **IOS**, **Android** and **Website of FPT Play**. With **21,000 comments from IOS** devices and **12,000 comments for both Android and Website**.



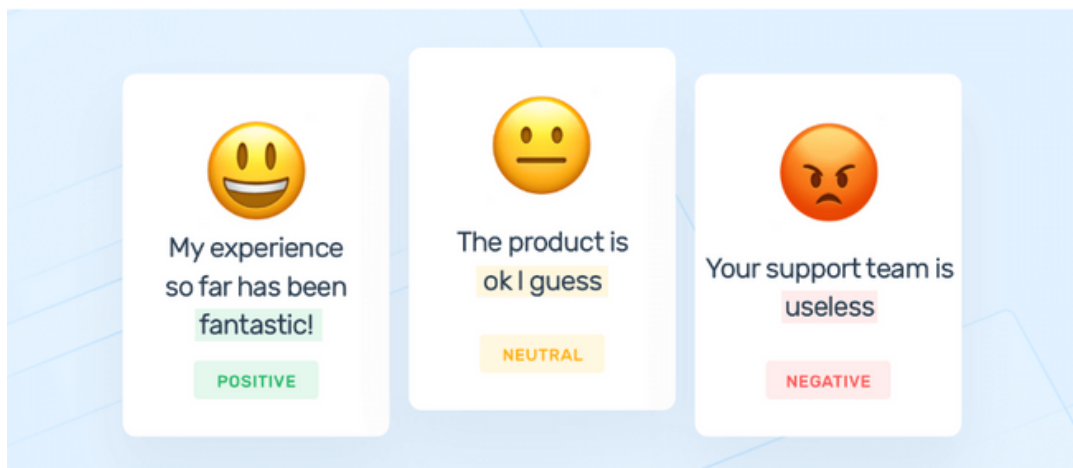
Top 10 most commented contents

- All of 10 contents that have most comments are in “Event” category.
- The most commented content was a livestream of Seagames 28 that has many Vietnamese athletes taking part in.

SENTIMENT ANALYTICS

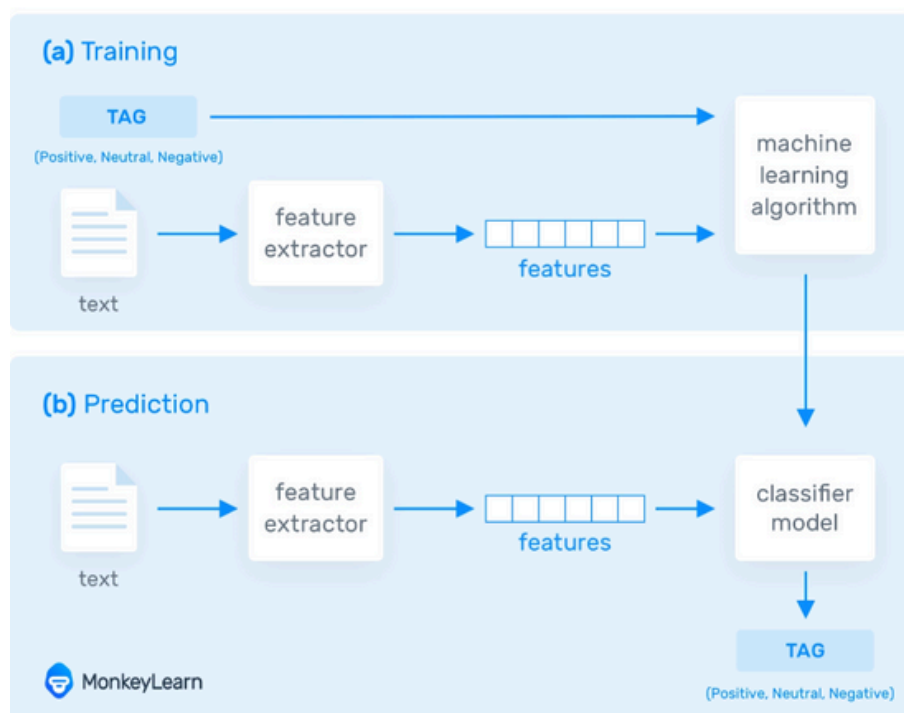
Introduction

Sentiment Analytics is the process of gathering the comments from the users to analyze the emotions of users while watching our contents. This Sentiment Analytics model run by the application of Machine Learning and Deep Learning into Neutral Language Processing (NLP)



Sentiment Analytics model take the text as an input, detecting positive and negative words in the text. And base on the machine learning algorithm, it can classify a text as positive, negative or neutral sentiment.

How Does Sentiment Analysis Work?

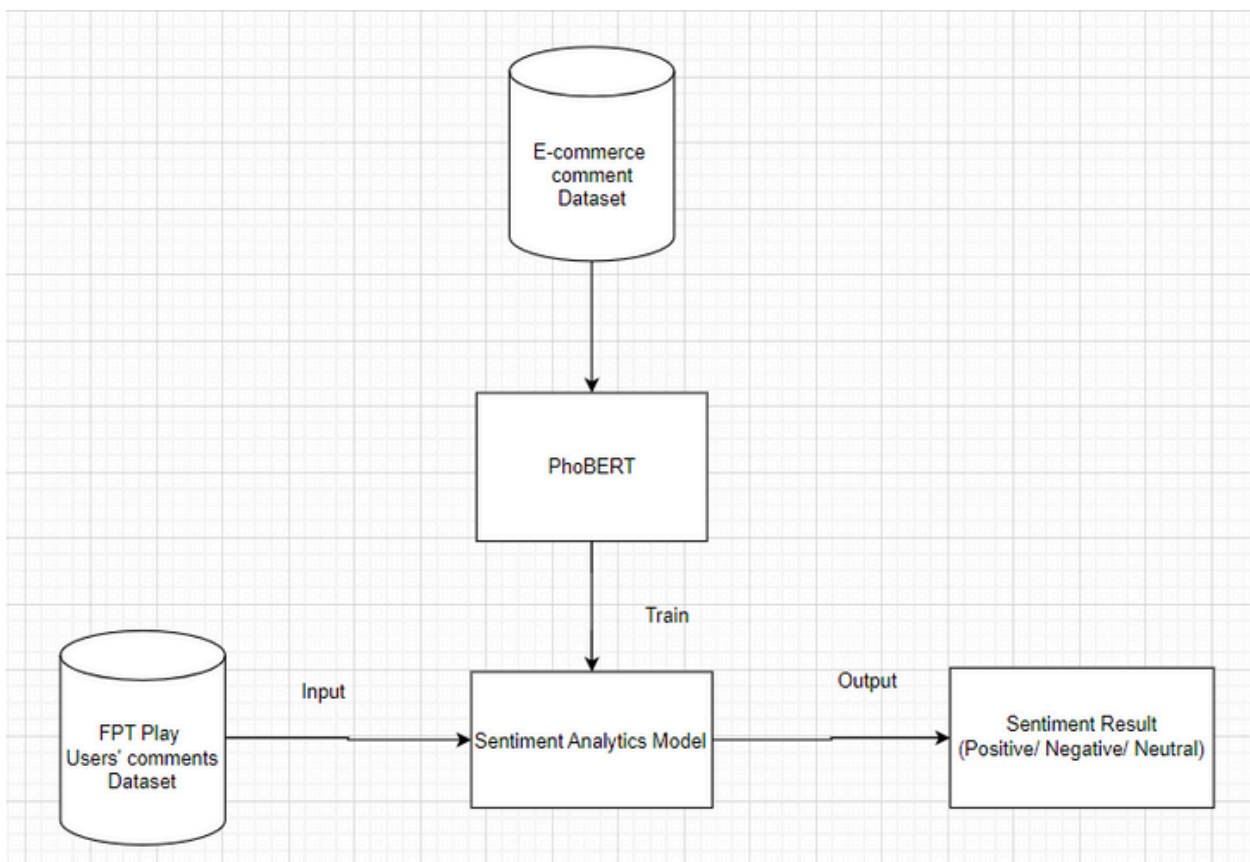


SENTIMENT ANALYTICS

Model Used

In this project, we use the pre-trained vietnamese analytics model which is programmed by VinAI named PhoBERT. And a sentiment analysis model is trained base on PhoBERT model and 30,000 comments of users on an E-commerce platform. This sentiment analysis model is written by author Pham Huu Quang.

Here is the code for reference: [PhoBERT](#), [PhoBERT base Vietnamese Sentiment](#)



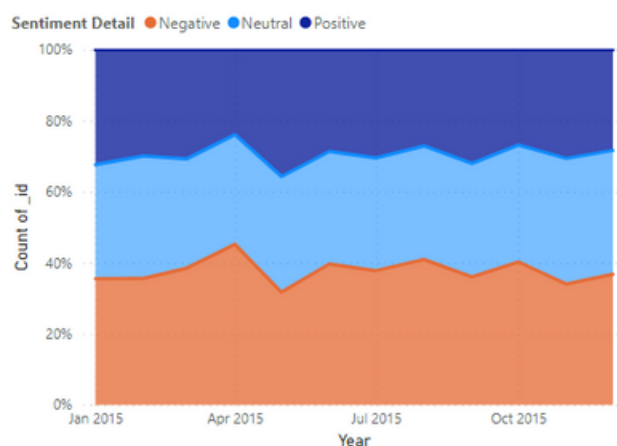
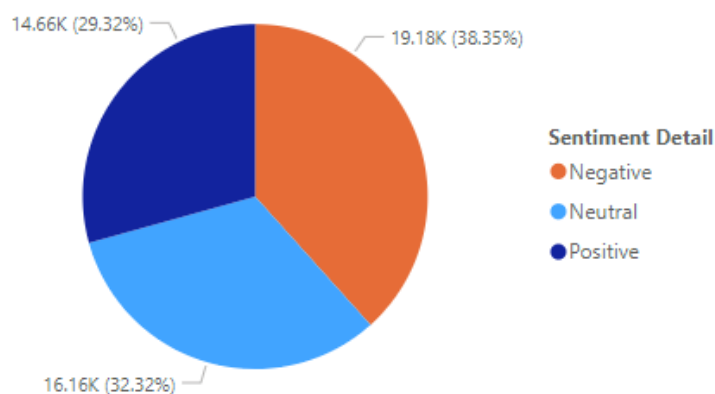
This diagram shows the process of analyzing sentiment of users on FPT play.

SENTIMENT ANALYTICS

Result

content	Sentiment Detail
	Positive
★★★★★	Positive
👍👍👍👍	Positive
dghjjkkvbbbb	Negative
sdffgghhhjj	Negative
dghklcgghjj	Positive
X. V mgyy	Neutral
Cac thanh c cho hiệp hai đi nhe	Positive
Bong chuyen nu viet nam gap u23thai lan	Negative
phim dôi v~	Positive
hay qua nhĩ	Neutral
Ok thôi ! Thu3 toan doan ma k dc thi thôi ve het đi	Negative
! v. "l okp k?m	Neutral
hay quá đi	Negative
Hiệp 2 bắt nhẽ u23	Positive
lcslx,lzms,	Neutral
tuyệt vời 4'08. lại phá KL rồi	Positive

After applying this sentiment analysis model into the comment dataset of FPT Play, this is the result that we can observe. The model can indicate which comments are positive, negative or neutral.



Overall Sentiment

Overall, the sentiment distribution is quite balance. However, negative comments still contribute a bit more than others, accounting for 38,35%. While Neutral has 32,32% and Positive is 29,32%. And the distribution overtime is also stable, there are no special behavior over the year 2015.

SENTIMENT ANALYTICS

Negative Comments Analysis

As presented in the overall chart, negative comments contribute the largest part in the comments of FPT play. Let's deep dive into it.

content	Sentiment Detail
Bong chuyen nu viet nam gap u23thai lan	Negative
Ok thôi ! Thu3 toan doan ma k dc thi thôi ve het di	Negative
hay quá đi	Negative
Bong chuyen nu viet nam u23thai lan	Negative
- nữ chính hk đẹp 🙄🙄🙄🙄🙄🙄🙄	Negative
:(nghĩ sao đã thế này hã vn ơi..mất mặt wa	Negative
- =)) Ông chúa nhìn chán đời vãi	Negative
2 0 cho messi	Negative
5k của tui	Negative
82 phút mà 0-3 thì gở thế nào đây . rõ to mồm , mạnh miệng thôi	Negative
A mặc áo trắng là ai nhờ	Negative
áo tung chao	Negative
Are 3 0 eve	Negative
- Ấy dà :v :v :V Naruto hay xen lẫn quá khứ quá. Hay thì hay nhưng hơi bị mất hứng :V :v :v	Negative
bài nhạc nền là gì vậy. cho em link, cảm ơn	Negative
Bạn Diễm : tập nhảy cầu ở bắc kinh hình như là 133 đó bạn	Negative
BKT co len	Negative
Bong da đang cap	Negative
buồn nắm VN ơi	Negative
- Cái này là trailer mà?	Negative
cập nhật xong, không xem được cái gì hết :(Negative
cha thay gì	Negative
chả thấy hài miễn bắc toàn hài miễn nam xem chán dữ	Negative
Chán nhĩ	Negative
chẳng xem dc	Negative
chelsea may man wa.	Negative
Có le 6-0	Negative
cố lên nào VN	Negative
co len thang 1-0 dc roi.	Negative
Coi may thang Việt Nam! Da chán	Negative

Most of the negative comments are about the sports, movies and the service of FPT Play. In terms of sports, it is easy to understand because watching sport give the viewers different or even contradict emotions. Same for movies, the negative comments can be written if the ciewers encounter uncomfortable experience. Such as advertising or pop-up while watching movies or the video quality is stable. There are also many negative comments about service of FPT Play such as: can not watch anything after update, videos is so lag, ...

=> Beside the negative comments that produced by contradict emotions due to characteristics of the contents such as sport or horror movies. We need to focus on the negative comments about our services and quality to improve our competitiveness on the market.

SENTIMENT ANALYTICS

Negative Comments Analysis

The bar chart below shows top 10 content that have highest percentage of negative comments where these contents have more than 100 comments.

object_id	% Negative	% Positive	Number of comments
55dc28b617dc13342bb1171d	0.61	0.09	191
55e7a49a17dc1350c51e58b8	0.56	0.17	878
55c7f84b17dc1344d4013099	0.55	0.16	101
5609088017dc130c36640a37	0.55	0.20	337
55b43f8217dc131ef3922efa	0.54	0.17	128
55dac54017dc1369c5a98d57	0.54	0.18	108
55ea4ad717dc1312a1a53c84	0.53	0.16	845
5617203f17dc135858f4949e	0.53	0.19	458
5611e24c17dc13642d47f017	0.52	0.21	450
55dd133217dc13342bb1174d	0.52	0.14	236

Top 10 contents that have most negative comments percentage

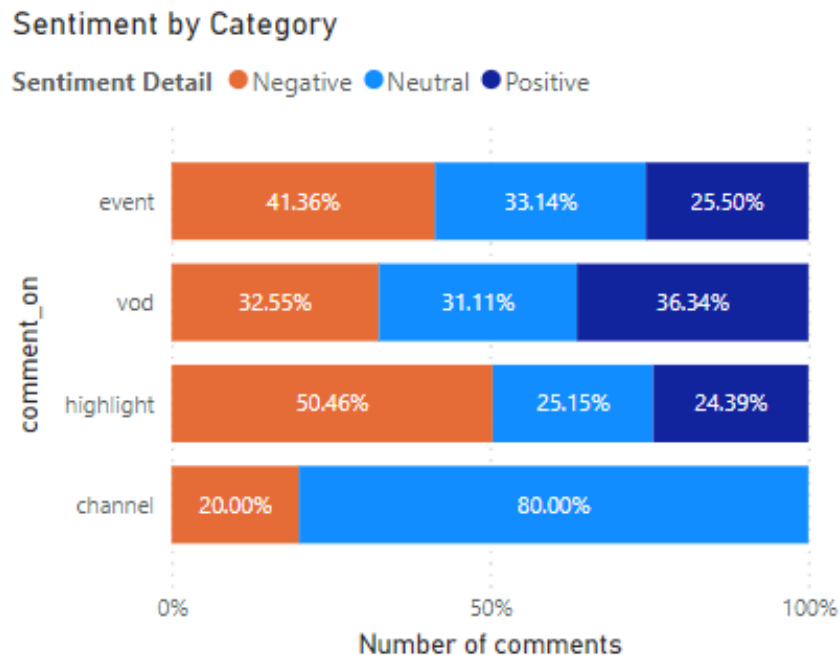
As can be seen in the chart, some contents have the proportion of negative comments upto 60%. Consequently, these contents have very low percentage of positive comments.

The common point of these contents is that they are all events or VODs about a contest or sport games or gameshows that have competitive characteristic such as: football match, The Voice Vietnam,...

=> **The competitive contents can leads to many negative comments from the viewers. Because different viewers will have different favourite teams or favourite candidates. Thus there will be mixed opinions among the viewers, and most of these opinions is negative.**

SENTIMENT ANALYTICS

Sentiment by Category



Sentiment distribution by Category

We can see that Highlight and Live Events have the highest percentage of negative comments. Events have a bit lower percentage of negative comments compare to Highlight, **41,36% and 50,46%**

=> **This is easy to understand because all the Highlight videos are relevant to sport and sport videos are easy generate mixed emotion and opition amongs viewers. While the live events can vary in content, including movies, music, gameshows, news, ...**

Other categories have much lower proportion of negative comments compare to two previous categories.

=> **This might be because of the characteristic of these contents. VODs and Channels' content are mostly movies, TV shows, comedies,... which have very few competitive characteristics.**

SENTIMENT ANALYTICS

Sentiment by Category

Now we will have a look at the content of the negative comments of these categories. The content of negative comments of Events and Highlight are mostly about sport and competitiveness contents. They mostly talking about football results.

content

0,02 điểm nhỏ wa
0.2 cho liverpool
0:1 tỷ số cuối cùng của trận đấu
0-0 bị đuổi 1 người
0-0 cho liverpool
0-0 lun bun ge
0-0 ok he gio
0-0 pháo gió mùa về r
0-0 thôi,ai bắt tỉ số
00 ui .read qua do ui
0-1 cho Mu
0-1 cho Việt Nam!
0-1 everton
0-1 là tỉ số cuối cùng của trận đấu MU và Liv ngày hôm nay
01202763261 núng bướm quá đi
0-1cho Doi ban
0-2 chủ nhà thua
0-3 cho mà xem
0-3 rồi còn đá gì nữa
0-3 thua rồi
0-3jQuery111107538321504835039_1437991269840?????????
0-4cho sen thai lan
0-5 là tỷ số đẹp!
0989681331 buồn quá zalo đi

Negative comments of Highlight and Events

SENTIMENT ANALYTICS

Sentiment by Category

On the other hand, the negative comments of VODs and Channels are about their bad experience with the platforms

ad ơi cho hỏi ko biết do may của mình hay là do fpt mà phim nào mình coi cũng đi dạt dạt qua trôi luôn
ad ơi fix dùm film này đi, ko play đv nha, dù đã coi rồi :v
ad ơi hỏng tập 16 r ko xem đc 🤔🤔🤔🤔🤔
ad ơi làm thuyết minh đi.... phim hay k có thuyết minh buồn lắm
ad ơi running man của hàn quốc sao lại xoá đi rồi ạ T.T
ad ơi s chưa có sub vậy a
ad ơi s coi TV cứ dc 5p hay 7p là mất hình thế ad !!!!
ad ơi s tập 5 coi k đc vậy bạn?
ad ơi sao các tập từ 2 đến 7 toàn là quảng cáo thế
ad ơi sao cứ báo ko phát được video thế
ad ơi sao cứ báo ko thể phát được video thế
ad ơi sao k úp tập 20 lên mà đã úp tập 21
ad ơi sao ko có tập 29 30 mà tập 27 28 xem bị lỗi ý
ad ơi sao tập 23 ko có tiếng v
ad ơi sao tập nào cũng có quảng cáo
ad ơi tập 3 không có sub việt
ad ơi up bản phụ đề đi mà :(nghe giọng ng thuyết minh chèn vô mà thấy nản :(diễn xuất của diễn viên đâu chỉ có ánh mắt, cử chỉ ko, còn cả giọng nói, ngữ điệu.... tuy không hiểu ngôn ngữ nhg vẫn cảm nhận đc mà :(quen xem phụ đề r bây h nghe tm k

Negative comments of VODs and Channels

=> Despite Highlight and Events have much more negative comments, we have to mind that the Negative comments in VODs and Channels are even more valuable to FPT Play. Because the Negative comments that related to the service qualities and users experience is very important to a service provider like FPT Play.

=> Furthermore, in terms of Highlight and Events, we have to have algorithms that segment out which comments are related to user experience and service quality such as commentators' knowledge, videos qualities, lagging stream to enhance the users experiences.