



Báo Cáo Phân Tích Dữ Liệu

FPT Play Sentiment

Giai đoạn báo cáo
2015 – 2022

Thực hiện bởi
Trang Do





MỤC LỤC

01 GIỚI THIỆU

02 PHƯƠNG PHÁP THỰC HIỆN

03 CÁC BƯỚC THỰC HIỆN

1. Xác định dữ liệu
2. Trực quan hóa và phân tích dữ liệu
 - a. Phân tích tổng quan
 - b. Phân tích cảm xúc (Sentiment) của người dùng

04 KẾT LUẬN VÀ ĐƯA RA GÓP Ý

Phần I.

Giới Thiệu

Công nghệ phát triển kéo theo nhu cầu giải trí ngày càng tăng. FPT Play ra đời để cung cấp nội dung hấp dẫn cho người xem. Phân tích bình luận sẽ giúp cải thiện chất lượng dịch vụ và thu hút khán giả.

Ngày nay, khi công nghệ ngày càng tân tiến thì nhu cầu giải trí của con người trở nên gia tăng nhanh chóng. Trong số đó chương trình truyền hình, phim bộ, phim hành động,...luôn là những lựa chọn hàng đầu sau ngày dài làm việc, học tập đầy áp lực.

Hiểu được điều đó, FPT đã cho ra đời trang web FPT Play - Nơi chứa đựng tất cả chương trình giải trí mà khách hàng mong muốn. Tuy nhiên, tại những chương trình này đều có rất nhiều bình luận hàng ngày và đa số đều không được thu thập để phân tích.

Đây có lẽ là một thiếu sót đáng được nhắc đến, bởi nếu có thể thu thập được loại dữ liệu này và rút ra insight khách hàng thì chúng ta sẽ hiểu thêm rất nhiều về khách hàng cũng như những trải nghiệm khi sử dụng dịch vụ của họ, từ đó đội ngũ cũng có thể vạch ra những chiến lược chăm sóc tốt hơn, đồng thời cải thiện chất lượng dịch vụ đáp ứng nhu cầu khách hàng.

Chính vì thế, bài phân tích những bình luận dưới đây sẽ giúp ban giám đốc, ban chiến lược đưa ra được những chương trình phù hợp và thu hút nhiều tầng lớp khán giả hơn.

Thực hiện bởi

Trang Do

Phê duyệt bởi

Tran Hoang Long

Phần II.

Phương Pháp Thực Hiện

Sentiment Analysis



Sentiment Analysis là một quy trình xử lý dữ liệu dạng văn bản nhằm xác định cảm xúc con người qua dữ liệu. Những cảm xúc này có thể đồng ý/tích cực, phản đối/tiêu cực hoặc trung tính.

Sentiment Analysis được thực hiện bằng cách áp dụng Machine Learning hoặc Deep Learning vào **NLP (Natural Language Processing)**

Chính vì thế, đây là công cụ mạnh mẽ giúp các doanh nghiệp hiểu sâu hơn về cảm xúc của khách hàng, từ đó tối ưu hóa trải nghiệm và đáp ứng tốt hơn nhu cầu của người dùng.

Mô hình PhoBERT



PhoBERT được nhiều người biết đến là một model phân tích ngôn ngữ Tiếng Việt hữu dụng phát triển bởi **VinAI**. **PhoBERT** được ứng dụng nhiều trong các mục đích như xác định loại từ, tự động hoàn thiện câu,...

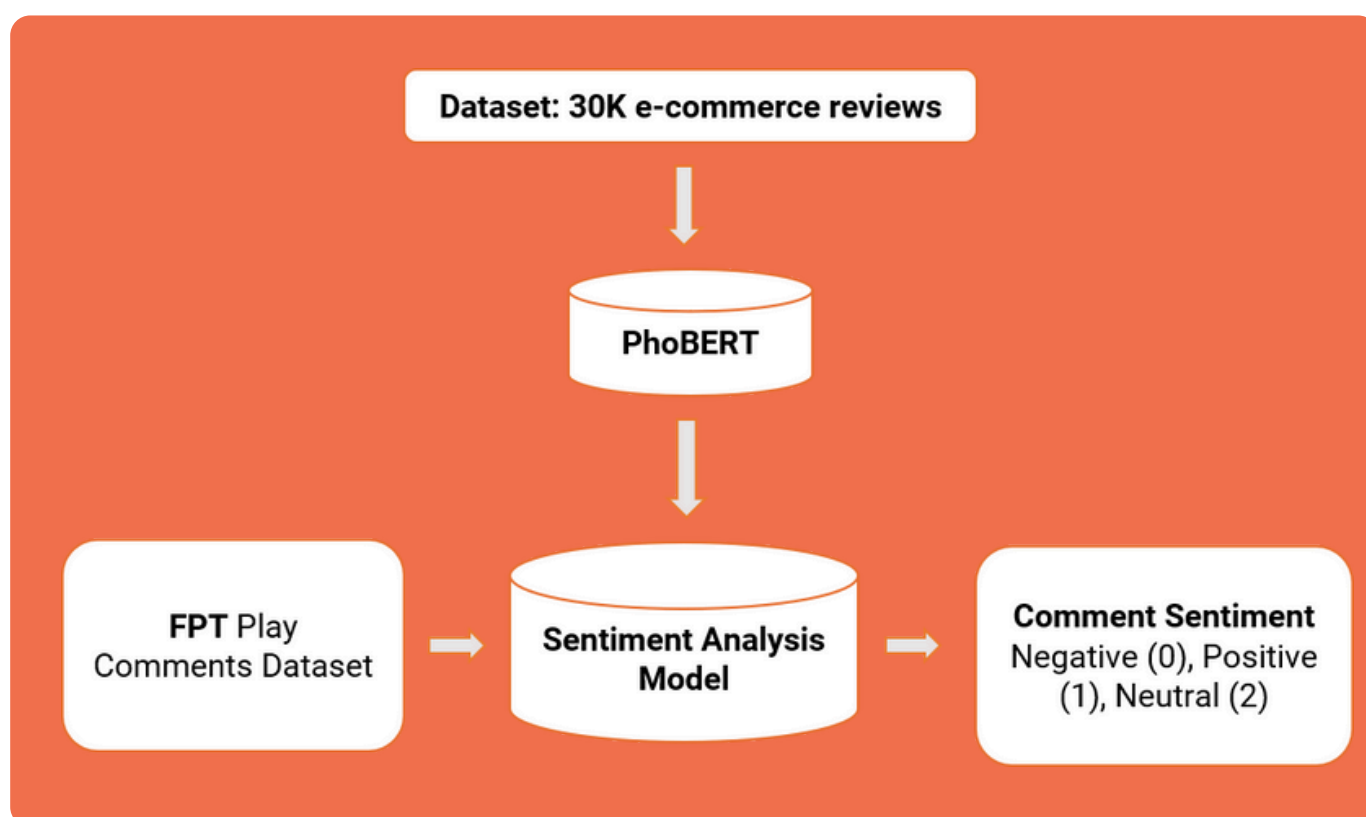
Để ứng dụng trong việc thực hiện Sentiment Analysis, chúng ta sẽ cần sử dụng 1 model đã được dựng sẵn, được train dựa vào model PhoBERT và data chứa tới **30.000 comment review** bằng tiếng việt trên một trang thương mại điện tử.

► Quy Trình Xử Lý

Quy trình xử lý để tạo ra kết quả của Sentiment Analysis được áp dụng trong dự án này như sau:

- 1** **Xử lý dataset:**
30.000 e-commerce reviews
- 2** **Áp dụng model** phân tích ngôn ngữ tiếng Việt PhoBERT
- 3** **Xử lý dữ liệu** FPT Play Comments Dataset với Sentiment Analysis
- 4** **Phân loại** cảm xúc từng comment: Negative (0), Positive (1), Neutral (2).

Sơ đồ cụ thể:



Tham khảo thêm về PhoBERT: *Pre-trained language models for Vietnamese*: [Link](#)

Tham khảo thêm về PhoBERT *Base Vietnamese Sentiment*: [Link](#)

Phần III.

Các Bước Thực Hiện

01 Xác Định Dữ Liệu

- Dữ liệu mô tả thông tin về các bình luận của người dùng trên nền tảng FPT Play được ghi lại trong khoảng 2015 – 2022.
- Dữ liệu được lấy từ MongoDB (NoSQL). Sau khi trích xuất dữ liệu thành công, tiến hành bước làm sạch dữ liệu với Python.
- Tổng cộng có 22 trường thông tin, tuy nhiên chỉ cần 8 trường thông tin, do đó cần loại bỏ các trường thông tin không cần thiết.

► Giải thích cột cần thiết

	Column name	Data type	Data sample	Meaning
1	_id	object	{'\$oid': '5553269a1a77ce5d65ea0604'}	Mã riêng của từng comment
2	user_id	object	464817	Mã riêng của từng user
3	comment_on	object	event	Thể loại của nội dung được comment
4	ip	object	42.116.8.164	Địa chỉ IP
5	object_id	object	54a53e4817dc131729ede7c8	Mã riêng của video
6	content	object	chuong trinh hay qua	Nội dung comment
7	timestamp	object	5/13/2015 10:32:00 AM	Thời gian comment được đăng
8	device	object	android	Thiết bị sử dụng để comment

► Sử dụng Python để clean và transform data

- Biến đổi trường thông tin '**_id**' để lấy mã
- Từ trường thông tin '**timestamp**' chúng ta sẽ sử dụng trường thông tin này để thêm và chuyển đổi thành 2 trường thông tin khác là '**date**' và '**hour**' để phân tích.
- Từ trường thông tin '**comment_on**' tạo thêm trường thông tin '**category**' và phân nhóm dữ liệu theo '**VOD**', '**Sport**', '**TV Show**'.

► Ứng dụng model PhoBERT Base Vietnamese Sentiment

Từ trường thông tin '**Content**' chứa nội dung comment của từng user, chạy PhoBERT Base Vietnamese Sentiment để phân nhóm các comment theo:

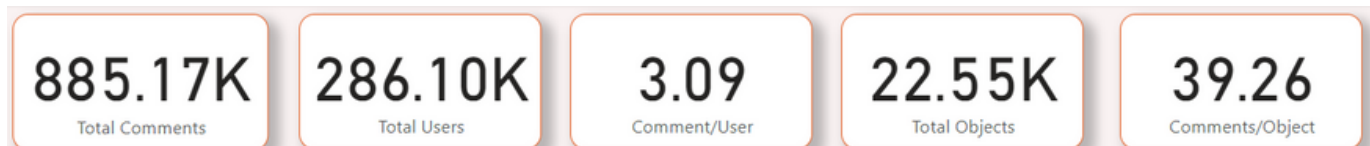
- **0: Negative** (Tiêu cực/phản đối)
- **1: Positive** (Tích cực/Đồng ý)
- **2: Neutral** (Trung tính)

► Kết quả sau khi xử lý 1000 bình luận:

```
sentiment_score
1      431
2      303
0      266
Name: count, dtype: int64
```

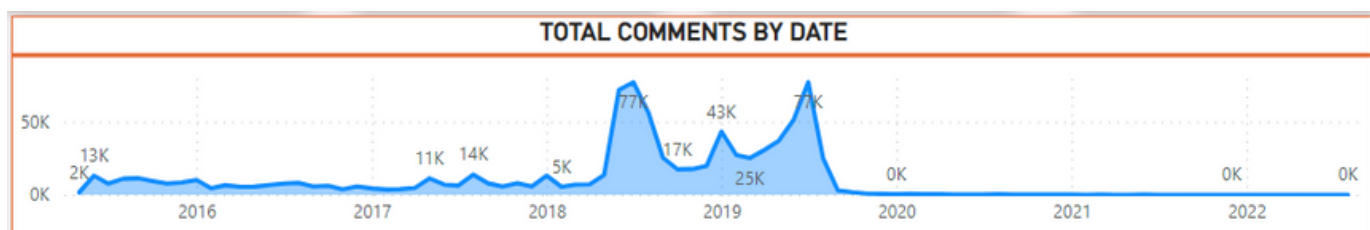
Phân Tích Tổng Quan

Dữ liệu cho chúng ta thấy được FPT Play sở hữu 885.17k bình luận của người dùng kể từ ngày 13/05/2015 đến ngày 30/08/2022. Chỉ ra trung bình mỗi user sẽ thực hiện 3.09 lượt bình luận/1 chương trình từ 3 thể loại chính là VOD, Sport và TV Show.

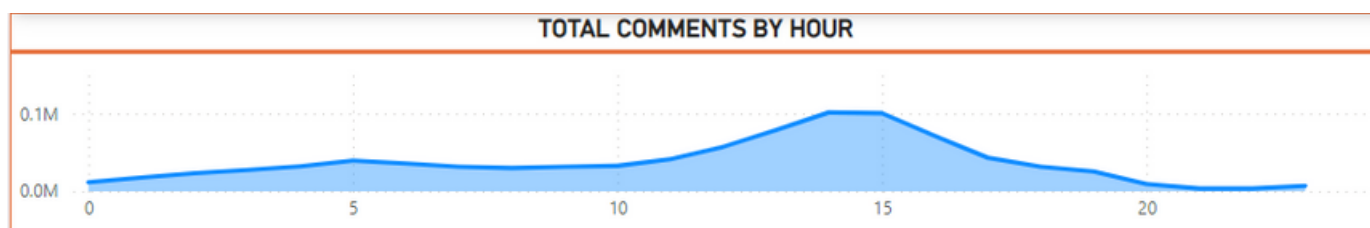


Chứng minh rằng không chỉ có số lượng lớn người sử dụng để xem nội dung mà người dùng cũng có sự tương tác tốt với FPT Play. Thể hiện rõ hơn ở chỉ số khi mà chỉ có **286.10K users** nhưng có đến hơn **885K lượt bình luận**.

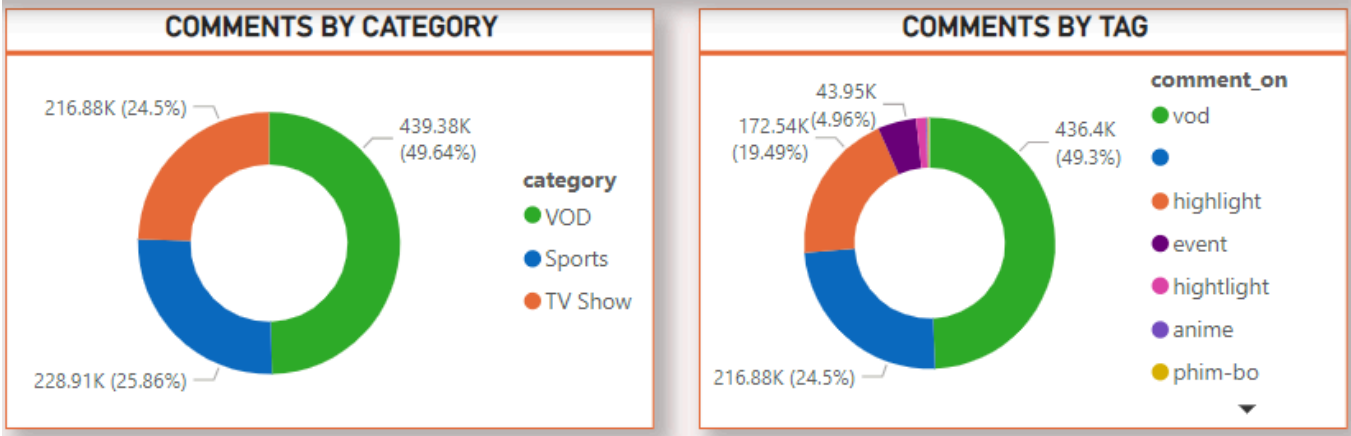
Tiếp theo, trong khoảng tháng 6,7,8 năm 2018 chứng kiến tăng mạnh lượt bình luận của người dùng khi đạt đỉnh với 77K comment. Bởi đây là thời điểm World Cup 2018 diễn ra. Theo sau đó là các tháng 6, 7 của năm 2019 khi lúc này diễn ra Vòng chung kết giải vô địch bóng đá Nam Mỹ với mức cao nhất cũng trong khoảng 77K comment.



Ngoài ra, lượt tương tác, bình luận của người dùng sẽ nhiều nhất vào khoảng từ 10 giờ - 16 giờ Việt Nam. Ngoài ra, lúc 00 giờ đến 5 giờ sáng dù rất sớm những vẫn có sự tương tác, có thể dễ hiểu khi mà gần như các giải đấu quan trọng thường rơi vào khung giờ này tại nước ta.

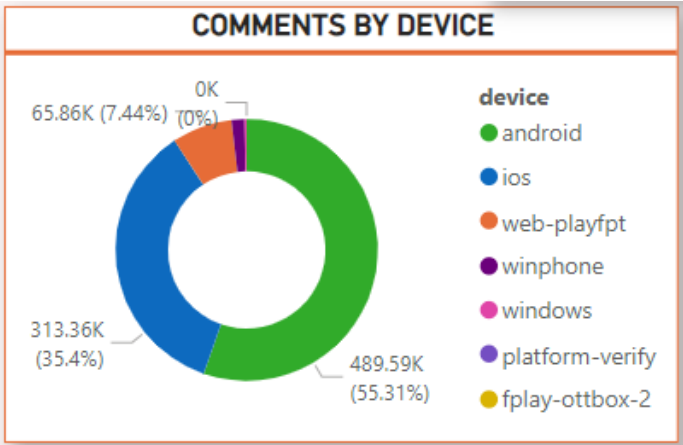


Thông thường, VOD sẽ sở hữu nhiều lượt bình luận nhất, sau đó là tiếp đến sẽ là Sport và TV Show. Có thể dễ hiểu, khi mà VOD gồm có những chương trình học online, phim bộ, phim hoạt hình,...thu hút rất nhiều lượng khán giả và fan hâm mộ. Xét về TAG, VOD vẫn tiếp tục chứng kiến số lượt bình luận nhiều nhất, sau đó là Highlight và Event. Trong khi đó, những nội dung tag khác như anime, anime, phim bộ, chương trình Tết thì lượng bình luận cũng không cao khi chỉ được xem và tương tác bởi một số user nhất định.



- Cho thấy nội dung của FPT Play quan tâm, đáp ứng kịp thời và chính xác thị hiếu của người dùng vào từng thời điểm.
- Hơn nữa, FPT Play cũng khá thành công trong xác định và quan tâm đến sở thích đa dạng của người tiêu dùng.

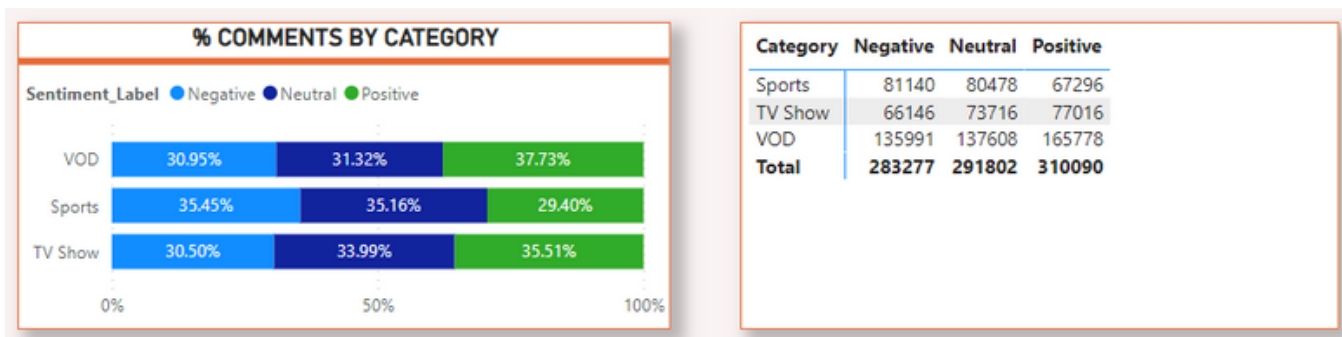
Tiếp đến, chúng ta sẽ đánh giá thiết bị sử dụng của người dùng. Qua biểu đồ, chúng ta có thể nhìn thấy tỉ lệ người sử dụng bình luận trên các thiết bị Android là phổ biến nhất, chiếm tới 55,31% so với các thiết bị còn lại khi mà ngoài điện thoại thì các TV thông minh đều có hỗ trợ cài đặt FPT Play từ Google Play Store. Tiếp theo sau đó là các thiết bị hệ IOS như Iphone chiếm tới 35,4%. Trong khi Website chỉ chiếm 7,44% trên tổng số lượt bình luận.



Cho thấy rằng, FPT Play luôn đem đến cho người dùng sự tiện lợi, gần gũi bởi ứng dụng App. Tuy nhiên, để tăng sự hài lòng của người dùng thì cần cải tiến nhiều hơn ở trải nghiệm website. Và FPT cũng nên tập trung nhiều hơn phân khúc khách hàng tại các tỉnh thành nhiều hơn.

► Phân tích cảm xúc (Sentiment) của người dùng

Trong số hơn 885k lượt comments trên các nền tảng của FPT Play thì trong đó sẽ có cả những bình luận tích cực, tiêu cực hay những bình luận trung tính.



Qua biểu đồ, ta thấy tỷ lệ bình luận tích cực, trung lập và tiêu cực đối với từng danh mục (Sports, TV Show, VOD) khá cân bằng. Không có một danh mục nào quá nổi trội về một cảm xúc cụ thể. Tuy nhiên, so với hai danh mục còn lại, VOD có tỷ lệ bình luận tích cực cao nhất, cho thấy người dùng hài lòng với nội dung của VOD hơn.

Sports:

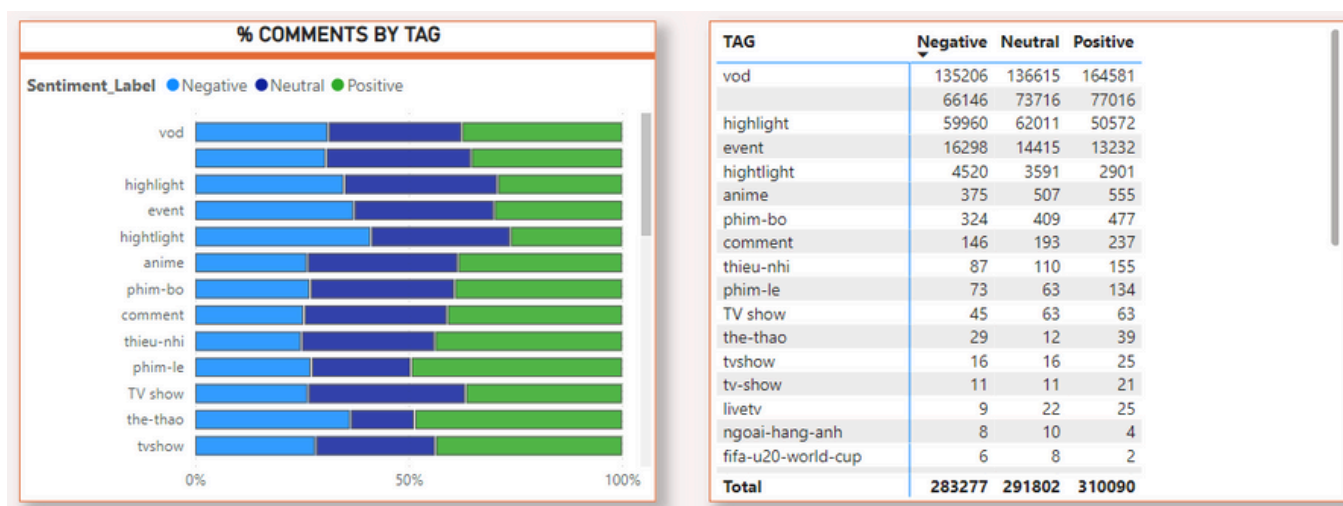
- Tỷ lệ bình luận tiêu cực và trung lập khá gần nhau, cho thấy người dùng có những ý kiến trái chiều về nội dung thể thao.
- Tỷ lệ bình luận tích cực thấp hơn so với VOD, có thể do yêu cầu của người xem về nội dung thể thao thường cao hơn.

TV Show:

- Phân bố cảm xúc tương đối đồng đều, cho thấy người dùng có nhiều đánh giá khác nhau về các chương trình truyền hình.

VOD:

- Như đã nói ở trên, VOD có tỷ lệ bình luận tích cực cao nhất, cho thấy người dùng hài lòng với nội dung video theo yêu cầu. Điều này có thể là do sự đa dạng về nội dung, chất lượng hình ảnh âm thanh tốt hay các tiện ích khác của dịch vụ VOD.



Tiếp đến, với biểu đồ tiếp theo thì nhìn chung các nội dung phổ biến như VOD, Highlight hay event đều có lượng tương tác vượt trội hơn so với những tag còn lại. Tuy nhiên, lượng tương tác cao cũng đi đôi với số comments tiêu cực cũng cao theo. Ngoài trừ anime khi lượng comments không nhiều nhưng comments tiêu cực cũng chiếm khá nhiều với 375 bình luận (26,1%).

Comment tiêu cực ở đây liên quan đến cả về nội dung như kết phim không đúng ý, phim ra chậm, đội đá thua, trọng tài đưa ra quyết định không hợp lý,...từ những người dùng phản ứng thái quá về cả dịch vụ, trải nghiệm như mạng lag, chậm, mất kết nối, quá nhiều quảng cáo, hay chất lượng âm thanh lồng tiếng không hay,...

Chính vì thế, để đảm bảo thu hút được nhiều người dùng, FPT Play cần phải **kiểm soát được những comment tiêu cực** về nội dung để có một văn hóa người xem văn minh hơn. Đồng thời, **cải thiện cơ sở hạ tầng** để đem đến trải nghiệm tốt hơn cho người sử dụng.

Phần IV.

Kết Luận Và Đưa Ra Góp Ý

Kết Luận

FPT Play đã thể hiện được sự quan tâm, đáp ứng nhu cầu, thị hiếu của người dùng với đa dạng các nội dung Sports, VOD, TV Shows, đặc biệt vào những dịp cao điểm trong tháng hè và xuân. Tuy nhiên FPT Play nên cải tiến để tăng trải nghiệm trên nền tảng website nhiều hơn khi các thiết bị android hay ios chiếm tỉ lệ quá vượt trội cũng như cần tập trung vào các khách hàng ở các tỉnh/thành phố khác ngoài những thành phố lớn. Các nội dung phổ biến xem nhiều đi đôi với lượng comments tiêu cực cao như vod, highlight, events cần giải quyết ở cả 2 khía cạnh gồm: dịch vụ và nội dung.

Góp ý

Các giai đoạn đầu năm và những tháng 6, 7, 8 chứng kiến lượng người xem nhiều hơn đi đôi với lượt tương tác, bình luận cũng nhiều hơn, bởi đây là khoảng thời gian nghỉ Tết, nghỉ hè của học sinh, sinh viên, người đi làm không có quá nhiều công việc dồn dập.

Chính vì thế, FPT nên:

- Cân đối nội dung để đáp ứng thị hiếu người tiêu dùng. Đồng thời, cân nhắc về các dịch vụ đi kèm khác như tích điểm đổi thưởng cho mỗi lượt tương tác, khuyến mãi combo sử dụng 2 đến 3 tháng,...
- Cân đối các chương trình quảng cáo, tiếp thị Marketing nhằm tăng trải nghiệm ở người xem
- Tăng băng thông mạng những lúc cao điểm
- Dồn nguồn lực để xử lý, lọc và nhắc nhở những bình luận tiêu cực liên tục. Từ đó, giúp tối ưu chi phí, tối đa doanh thu và thu hút người tiêu dùng.



Trang Do

Data Analyst

Contact

036 319 48 81



dothitrang691999@gmail.com



linkedin.com/in/dotrang0609

