

|  |  |
| --- | --- |
| *FPT Play* Sentiment Analysis | |
|  |  |
|  |
|  |
| Anh Nguyen Qualitative - Comments Analysis  2024 |  |

### GIỚI THIỆU

##### Tổng quan

Báo cáo này phân tích về bình luận của người dùng FPT Play trong giai đoạn 2015-2022.

Bảng sau giải thích về dữ liệu thu thập được:

A screenshot of a computer

Description automatically generated

1. Truy cập vào Data lake (ở đây sử dụng Mongodb)
2. Đọc dữ liệu và phân tích bằng mô hình máy học PhoBERT (machine learning by Python) được phát triển bởi VinAI: Các mô hình ngôn ngữ tiền huấn luyện cho tiếng Việt, sử dụng các reviews trên một trang thương mại điện tử.

A diagram of a diagram of a diagram

Description automatically generated

Giải thích về Sentiment analysis: Phân tích quan điểm là một ứng dụng của trí tuệ nhân tạo, nó sử dụng các thuật toán phức tạp để xử lý ngôn ngữ tự nhiên của con người (NLP) và xác định các đặc điểm cảm xúc tiêu cực/tích cực tại một thời điểm thông qua văn bản hoặc lời nói.

A screenshot of a phone

Description automatically generated

A screenshot of a computer

Description automatically generated

### Kết Quả

A computer screen shot of a program code

Description automatically generated

Chúng ta test mô hình với một câu như “lồng tiếng mà như tập đọc nghe không nuốt nổi”, kết quả sẽ cho 3 loại là NEG, POS, NEU, với % NEG cao, câu này mang sắc thái tiêu cực.

A graph of different colors

Description automatically generated with medium confidence

Có thể thấy rằng bình luận tích cực thấp hơn bình luận tiêu cực trong kết quả được tìm thấy.

### Phân Tích

A number on a white background

Description automatically generated

Mùa cao điểm của user vào năm 2018, các trận đấu bóng đá World Cup đang diễn ra, có thể thấy được sự khác biệt trong biểu đồ lưu lượng bình luận.

A graph with numbers and a line

Description automatically generated

Trong tất cả các kênh đang được phát sóng trên FPT Play thì Học Online và Chương trình Tết có mức độ bình luận “tích cực” cao nhất. Thể thao có lượt bình luận “tiêu cực” cao nhất, sau đó là đến sự kiện. Điều này khá dễ hiểu khi thể thao là thị trường tạo nên nhiều cảm xúc và thị trường thể thao tại Việt Nam thường có cái nhìn khá bi quan.

Trong khi nội dung “Film” nhận về số lượt bình luận cao nhất, tiếp theo sau đó là VOD và Sự kiện.

A screenshot of a graph

Description automatically generated A screenshot of a computer

Description automatically generated

Thiết bị được sử dụng nhiều nhất là Andoird, chiếm đến hơn 55,3% , cần sử dụng chiến thuận marketing ads hợp lý để target vào các thiết bị khác nhằm tăng user.

A pie chart with text

Description automatically generated

### 