FPT Play

SENTIMENT ANALYSIS REPORT

DATE :
15 July 2024

PREPARED BY :
Thu Hien

TABLE OF CONTENT

ABOUT FPT PLAY 01

PROJECT OVERVIEW 02

METHODOLOGY 03

DASHBOARD OVERVIEW 04

ASSESSING CUSTOMER SATISFACTION 05

ABOUT FPT PLAY

FPT Play is a Vietnamese television and digital content brand, owned by FPT Telecommunications Joint Stock Company (FPT Telecom) under FPT Group. This is a popular on-demand and fee-based online television service in Vietnam besides MyTV, Clip TV... FPT Play application can be viewed on many devices such as smart TVs, smartphones, tablets, Android TV Box or websites. Besides domestic and international copyrighted content, FPT Play also participates in producing exclusive movies, entertainment and sports programs.

FPT Play owns nearly 200 domestic and international television channels. In addition, FPT Play content store also provides online learning applications, home sports training, online Event applications... allowing users to experience many needs: learning, entertainment, play games... on multiple device platforms.

FPT Play's mobile application version was first launched on the app store in May 2013. At that time, cable TV and Internet service providers were still in a monopoly position in providing services such as entertainment content, videos, etc..

PROJECT OVERVIEW

OBJECTIVE

- Sentiment analysis (or opinion mining) is a natural language processing (NLP) technique used to determine whether data is positive, negative or neutral. Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand customer needs.

- In this project, analyzing customer reviews, comments, and feedback help understand the sentiment behind them and identify areas (programs, channels, events,...) for improvement and addressing customer concerns, ultimately enhancing customer satisfaction.

DATASET

Colume name	Define	Information
_id	Id of the comment	not necessary
ranking	This could represent the ranking or score assigned to the comment. Since it is 0, it might indicate no ranking or a default	All records are 0 => not necessary
layer	This might indicate the depth or level of the comment in a threaded discussion or nested comments structure. A value of	All records are 0 => not necessary
user_id	The unique identifier for the user who made the comment.	Indentify user
like	The number of likes or upvotes the comment has received.	Level of agreement comment receive
user_email	The email address of the user who posted the comment.	Indentify user
comment_on	This could indicate the context or the item on which the comment was made. It is empty here, so it might be a	Name of programe user comment on
ip	The IP address from which the comment was posted. This can be used for logging or moderation purposes.	Find out province in which user live
user_fullname	The full name or username of the user who posted the comment	username
publish_status	The status of the comment, indicating whether it is published or not. A value of 1 typically means it is published.	whether Comment is published or not
object_id	The unique identifier of the object (e.g., video, article, product) to which the comment pertains.	not necessary
content	The actual text content of the comment.	content of the comment
reviewer	This could indicate the user or admin who reviewed the commer	not necessary
report_count	The number of times the comment has been reported by users.	All records are 0 => not necessary
review_status	The status of the review process for the comment. A value of 0 could indicate it has not been reviewed.	not necessary
timestamp	The timestamp of when the comment was posted, usually in Unix epoch time (seconds since January 1, 1970).	Convert to datetime
device	The type of device from which the comment was posted.	Device user is usint to post comment
report_reason	This could list the reasons why the comment was reported, if an	All records are null => not necessary
dislike	The number of dislikes or downvotes the comment has received	All records are 0 => not necessary
comment_status	This could indicate the current status of the comment (e.g., acti	not necessary
device_id	The unique identifier of the device from which the comment was posted. This can be used for tracking or moderation	not necessary

The original data is a JSON file, contains 885,169 elements and is stored in MongoDB (NoSQL)

Since my system configuration is not high-end, I just retrieve 30,000 records for analyzing.

After successfully extracting data from MongoDB , I use python to clean and transform data:

- There is a total of 21 columns, however just some of them are necessary, so unnecessary columns need to be removed.
- Transform column “ip” to column 'province' using geoip2.database to get information about province in which user lives. For example: 116.105.57.67 => Ho Chi Minh City.
- For the “timestamp” column, I convert this to “date_time” column for analysis. For example: 1628960400 => 15/08/2021.

METHODOLOGY

For efficient sentiment analysis, first of all, I remove irrelevant information (e.g., HTML tags, special characters) from “content” column.

As data is in Vietnamese, I use the ***PhoBERT model*** for sentiment analysis.

The PhoBERT model is a RoBERTa-based model that is pre-trained on a large corpus of Vietnamese text. It is available in the transformers library.

PhoBERT is a strong model for Vietnamese text, but it was trained on text with full diacritics. Then it is well-suited for Vietnamese text with accents. For text without accents, I pre-process the data to match the model's input requirements with ***Transformer model for inserting Vietnamese accent marks***. This model inserts accent marks (diacritics) for Vietnamese texts that don't have them (or texts with some words accented and some not).

Link of PhoBERT model:

https://pdfs.semanticscholar.org/74fc/832dd6c77253595cf3c1c852045c8da93c13.pdf

Link of Transformer model for inserting Vietnamese accent marks:

https://huggingface.co/peterhung/vietnamese-accent-marker-xlm-roberta

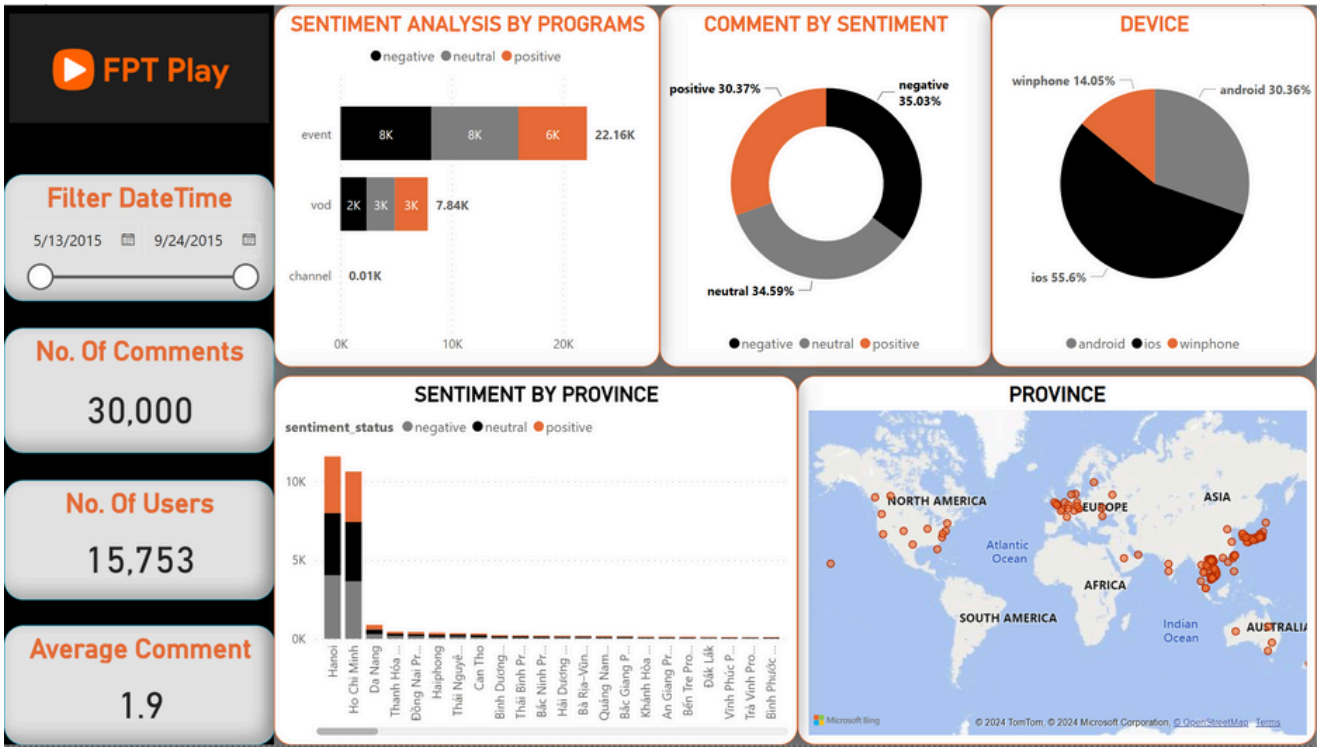
To optimize the Accent Marker Model and PhoBERT model for faster inference:

1. Batch Processing: Instead of processing each comment one by one, I can process them in batches. Batch processing allows me to process multiple comments at once.
2. Move to GPU: leveraging a GPU help me significantly speed up inference.

After transforming data and implementing sentiment analysis using python, I load analytical data into SQL sever:

	user_id	user_email	user_fullname	province	date_time	comment_on	content	accented_content	sentiment	publish_status	comment_status	likes	device
1	464817	13@fpt.com.vn	tuyen tuyen	Ho Chi Minh	2015-05-13 10:32:17	event	chuong trinh hay qua	chương trình hay quá	1	1	0	12	android
2	28	lokiheero@gmail.com	Quốc Trung	Ho Chi Minh	2015-05-13 10:35:37	vod	hehe không biết chừng nào có tập 6	hehe không biết chừng nào có tập 6	1	1	0	4	android
3	3839	xxx@gmail.com	xxxxxumuyu	Ho Chi Minh	2015-05-14 01:23:52	vod	phim hay qua	phim hay quá	1	1	0	4	android
4	527201	cloudytang@gmail.com	Đàm Ngọc Vân	Ho Chi Minh	2015-05-14 03:20:23	vod	hay qua do	hãy quá đỗ	0	1	0	5	android
5	527201	cloudytang@gmail.com	Đàm Ngọc Vân	Ho Chi Minh	2015-05-14 03:21:02	vod	comment cai	comment cái	2	0	2	0	android
6	464817	13@fpt.com.vn	tuyen tuyen	Ho Chi Minh	2015-05-14 03:28:14	vod	bóc tem 8kyu	bóc tem 8kyu	2	0	2	1	winphone
7	16955	nguyenducmanh609@gmail.com	nguyenducmanh609@gmail.com	Đồng Nai Province	2015-05-14 05:47:39	vod	phim hay lắm tải về thế nào ta	phim hay lắm tải về thế nào ta	1	0	2	0	android
8	16955	nguyenducmanh609@gmail.com	nguyenducmanh609@gmail.com	Đồng Nai Province	2015-05-14 05:47:44	vod	phim hay lắm tải về thế nào ta	phim hay lắm tải về thế nào ta	1	1	0	2	android
9	277615	pexuxu8@gmail.com	xu xu	Ho Chi Minh	2015-05-14 06:33:38	vod	dễ thương ghê	dễ thương ghê	1	1	0	7	android
10	464817	13@fpt.com.vn	tuyen tuyen	Ho Chi Minh	2015-05-14 06:45:16	vod	jsggsgsdhdhgd	jsggsgsdhdhgd	1	0	2	1	winphone
11	132514	vta94vta@gmail.com	Vũ Tuấn Anh	Lâm Đồng Province	2015-05-14 07:15:30	vod	ko ai ns nâng gì à	ko ai ns nâng gì à	0	1	0	4	android
12	130	gian.phan@gmail.com	Phan Thanh Gian	Ho Chi Minh	2015-05-14 08:01:52	vod	hello hay qua	hello hay quá	2	1	0	6	android

DASHBOARD OVERVIEW

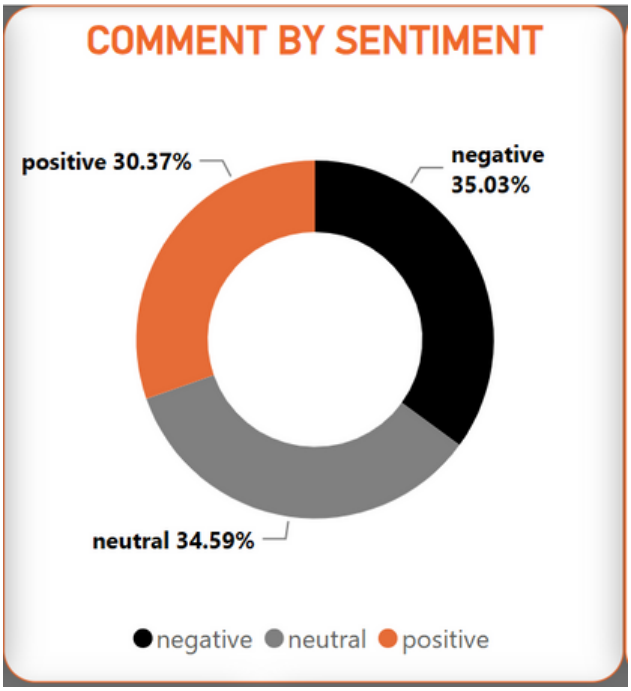


DASHBOARD OVERVIEW

Some key information from Dashboard:

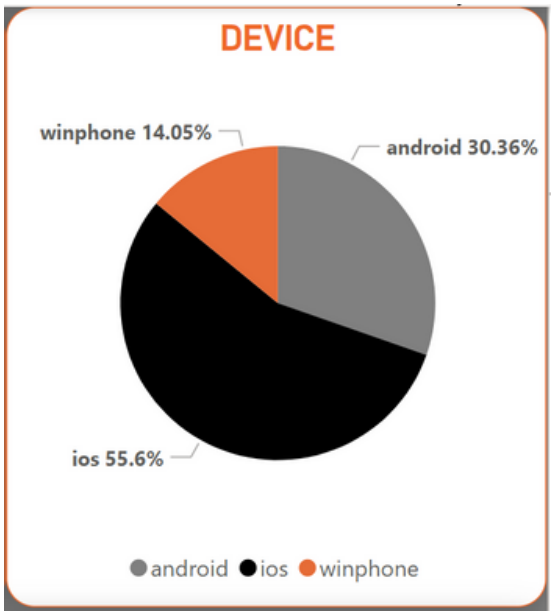
- There were **30,000 comments** in period of 130 days from 5/13/2015 to 9/24/2015.
- There were **15,753 users** who left comment, resulting in an **average comment** is **1.9**. This means that each interactive user left 2 comments on FPT Play.
- Almost users who use FPT Play live in **Ha Noi and Ho Chi Minh City.**
- FPT Play is not only popular in Vietnam but also in some European countries, America, Australia and other Asian countries.
- The amount of positive comments, negative comments and neutral comments **is relatively equal.**
- More than half of FPT Play users use the **iPhone operating system (iOS).**

ASSESSING CUSTOMER SATISFACTION WHEN USING FPT PLAY THROUGH SENTIMENT ANALYSIS

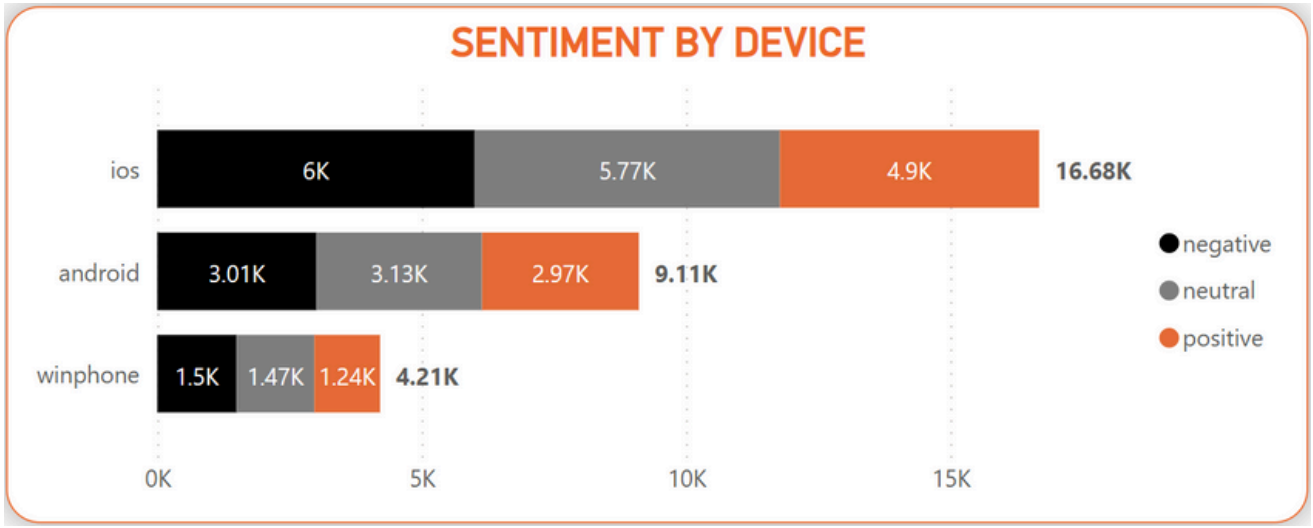


- The sentiment distribution is fairly balanced, with no single sentiment overwhelmingly dominating the others.
- Users have mixed experiences, with a slight lean towards negativity.
- Insight for Service: negative feedback need to be addressed, positive experiences need to be encouraged, and neutral users need to be engaged in ways that could turn their sentiment more positive.

SENTIMENT ANALYSIS BY DEVICE



- From FPT Play mobile application version was first launched in 5/2013 to 9/2015, users essentially watch on mobile application.
- The amount of FPT Play users who use iOS is more than the sum of those who use winphone and android.
- It seems that iOS users tend to leave more comments than Android and winphone users.



- Positive comments are the lowest across all devices but are fairly balanced in comparison to negative and neutral comments.
- This chart show that users experience are the same when they use iOS, android or winphone.
- Although some experiences might be similar (e.g., the presence of positive, negative and neutral sentiments across 3 platforms), there are notable variations in user engagement and satisfaction depending on the device used, based on the total of comment on device.

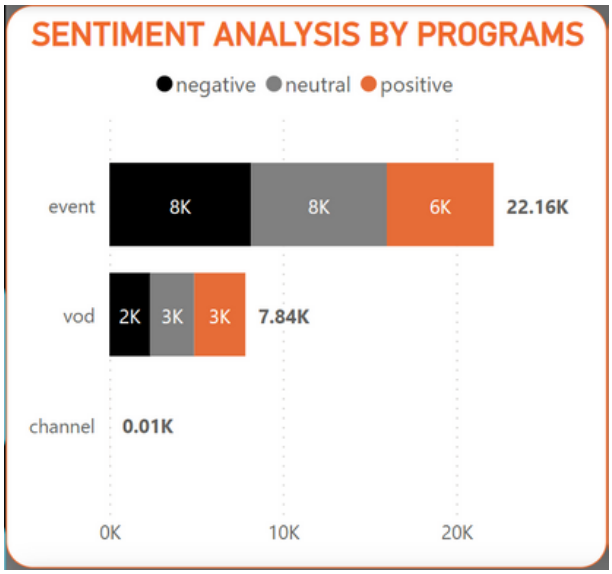
SENTIMENT ANALYSIS BY PROGRAMS

FPT Play offers various types of content:

1. VOD (Video on Demand): pre-recorded content that you can watch at any time (such as movies, TV shows, series, documentaries), as opposed to live broadcasts.

2. Channel: the live TV channels that FPT Play streams. These are similar to traditional television channels. Users can watch these channels live as they are being broadcast.

3. Event: live or scheduled broadcasts of specific events on FPT Play (ports matches, concerts, live shows, or other special programs). Events on FPT Play are often promoted ahead of time, and users can tune in to watch them live or sometimes catch a replay later.

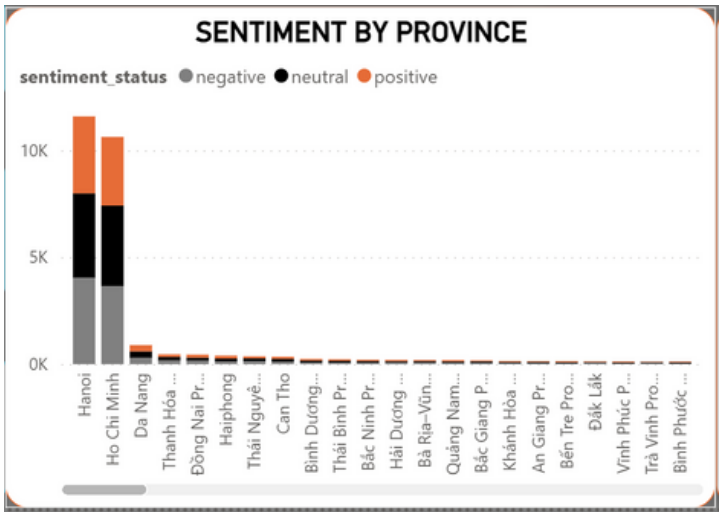


- Almost user's feedbacks are on event.
- Users' feedback on event is three times higher than the sum of VOD (Video on Demand) and channel feedbacks.
- Channel is the least interest program.

=> Users are more likely to be motivated to share their thoughts on something special or new (Event is unique, immediate, timely, communal, and emotionally engaging nature, meanwhile VOD and Channel is routine and predictable).

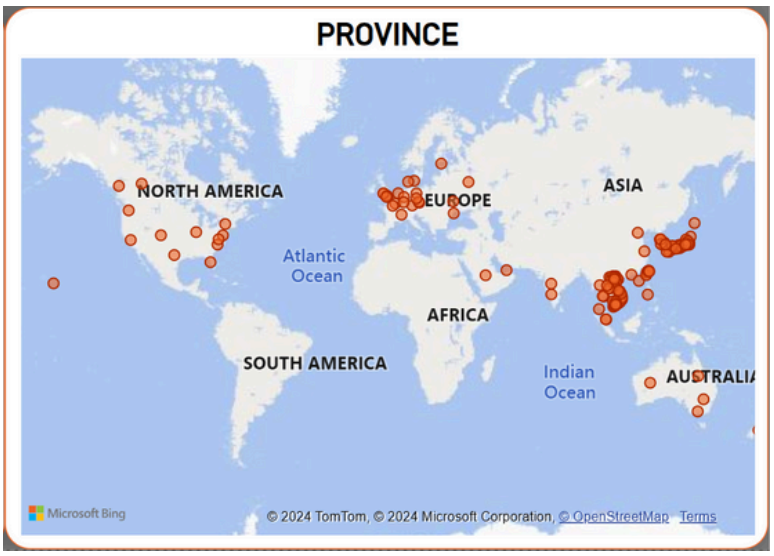
=> This insight could be leveraged to enhance engagement strategies, particularly by creating more event-like experiences or by finding ways to bring similar excitement to VOD and channel programs.

SENTIMENT ANALYSIS BY AREA



- Almost users concentrated in Ha Noi and Ho Chi Minh city.
- Hanoi and Ho Chi Minh City: The distribution is similar, with a strong presence of positive sentiment, followed by negative and neutral.
- The remaining provinces have significantly fewer comments.

- The map shows that the majority of user engagement is concentrated in Asia (Vietnam and Japan are two country have the largest amount of users), followed by North America and Europe.
- These regions are likely the primary markets or areas of interest.



Insight for Emerging Markets: Australia and some regions in South America and Africa could be considered emerging markets. While there is some activity, these areas may have potential for growth if targeted effectively.

Insight for Global Strategy: The map indicates that the FPT Play has successfully reached users globally, but there is room to expand in less represented regions like South America and Africa.

THANK YOU

Source code: ***FPT Play Sentiment Analysis Project***

Thuhien1993.ueh@gmail.com
0986.844.911