



Virtual Event 15-18 June
2020
**2020 Asia-Pacific
Statistics Week**

Leaving no one and nowhere behind

Listening the public opinion? An approach from big data with the case of revision GDP in the period 2010-2017 in Vietnam

Action Area B: **Assuring quality and instilling trust in statistics (SB2)**

Presenter:

Nguyen The Hung

The General Statistics Office of Vietnam

#apstatsweek2020



#apstatsweek2020



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Introduction

- ✓ Importance of Social listening in management at all levels
- ✓ Listening to the opinions of people helps us obtain clear evidences in the real situation
 - Traditional research is biased and expensive
 - Development of computer science & advancement of data analysis => Management now have more solutions.



#apstatsweek2020



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Introduction

- ✓ Using **sentiment analysis** through **Social Listening Tool** is one of the key transformation, provide numerous insights for managers
- ✓ Using a Social Listening Tool to collect social views on specific event and product;
- ✓ Applying Big Data through social media for sentiment analysis
- ✓ Determining the social response in term of sentiments, and user's basic information



#apstatsweek2020



Virtual Event 15-18 June 2020

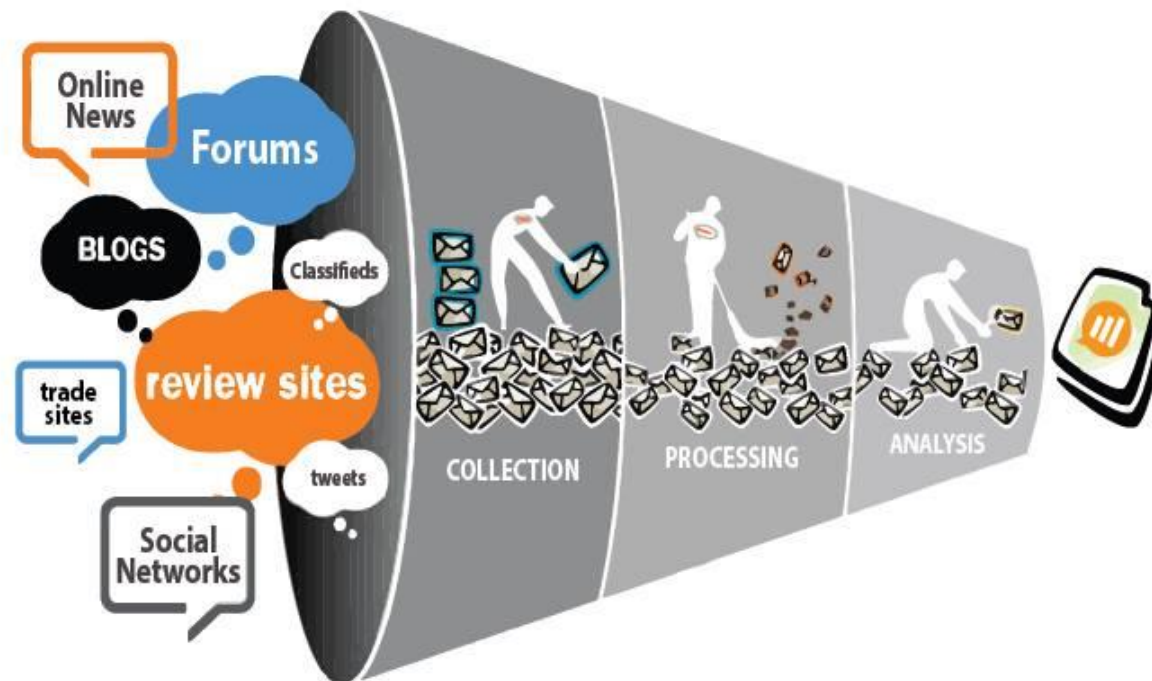
2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Methodology

In 2019, Vietnam has revised the size of GDP in the 2010-2017 period

- doubt the accuracy of the data as well as questioning the independence nature of the statistical office
- Others support the reevaluation to get a more accurate picture of the economy.





Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Data Collection

The social listening tool is R – product, self developed

Range time from 2019, 16st August (the first time the GSO launched the project revising GDP) to 2020, 31th January (01 month after GSO published the result)

Data source: (1) website; (2) personal post; (3) fan page post; (4) group post; (5) comments; (6) forum and (7) YouTube.

It is larger source than Google with searching only on website.



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Data processing

- ✓ Natural Language Processing (NLP)
 - ✓ Machine learning
 - ✓ Deep learning
-
- Positive sentiment is the content of the mention supporting for this issue;
 - Negative sentiment is the content of the mention opposing to this issue;
 - Neutral is the content of the mention is not closely related to the research problem. In other words, it could be considered as noise.



#apstatsweek2020



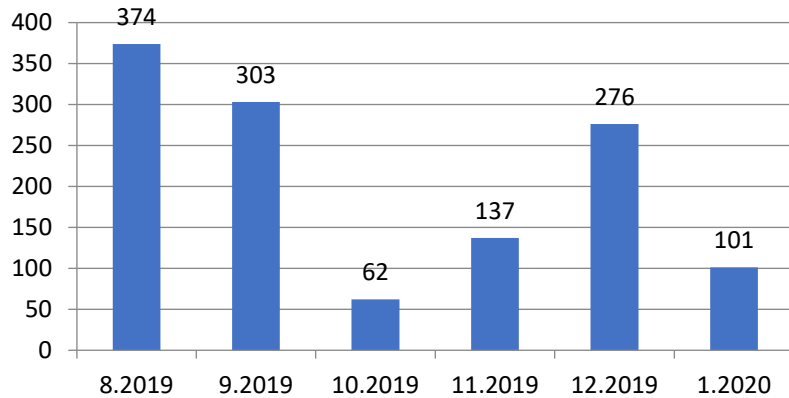
Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

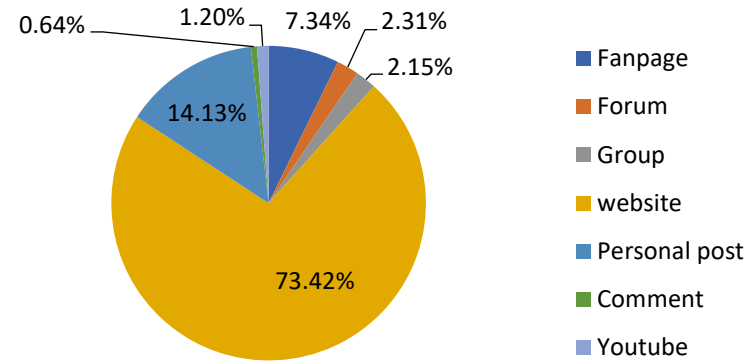
Leaving no one and nowhere behind

Result

There are 1.253 mentions about the revising GDP



The number of mention by month



Sources of mentions



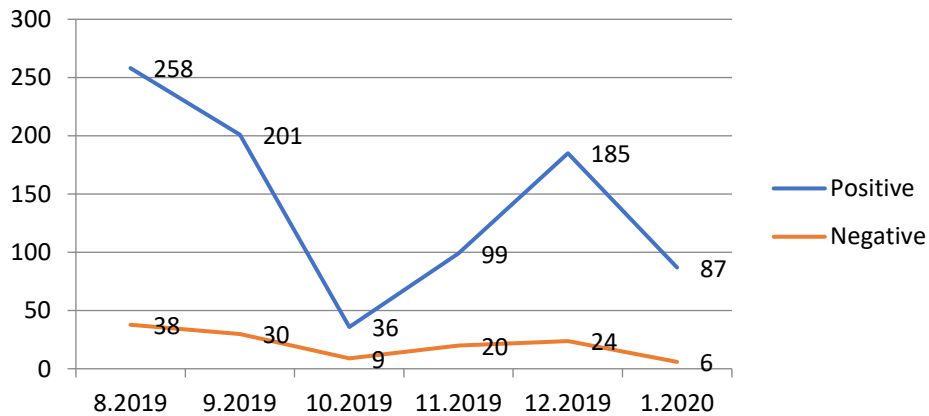


Virtual Event 15-18 June 2020

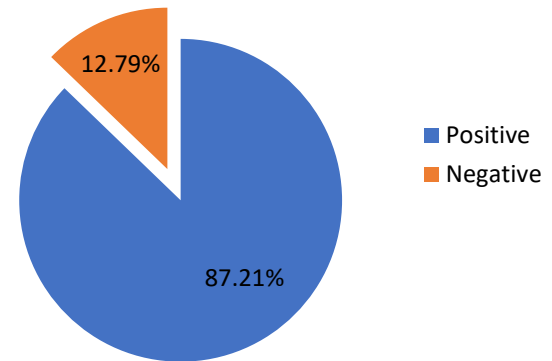
2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Result



The proportion of sentiments from December, 2018 to September, 2019



The proportion overall of sentiment



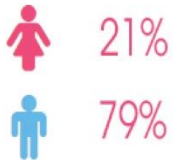
Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

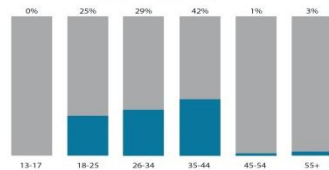
Leaving no one and nowhere behind

Result

DISTRIBUTION BY GENDER



POPULAR AGE RANGE



GENDER DISTRIBUTION BY AGE



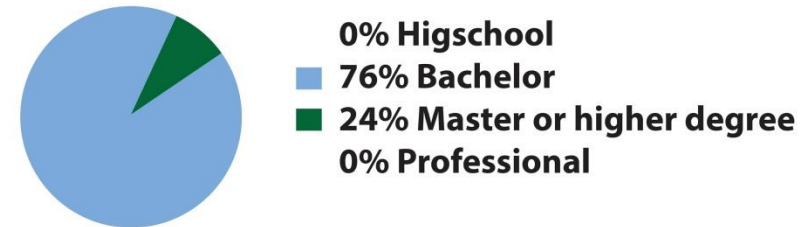
Gender distribution of subscribers

DISTRIBUTION BY LOCATION AND EDUCATION

DISTRIBUTION



EDUCATION



Location and education distribution of subscribers





Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Discussion, Conclusion and Recommendations

Revising the size of GDP is a regular task of national statistical offices but this task can cause public misunderstandings about the objectivity of the statistical data.

To minimize negative options that cause instability in the society, reduce trust in statistical products, the propaganda needs to be carried out through official communication channels such as GSO website, mass media and special need to focus on social networks

The propagation and exchange idea also need to be carried out through workshops, seminars at universities, research institutions because this is a place for people with deep expertise and have a great influence in society.

These activities should focus on taking place around the time of announcing of the implementation plan and publishing the results



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Discussion, Conclusion and Recommendations

The social listening tool has excellent advantages compared to conventional surveys, such as:

- (1) The number of observations is much greater
- (2) Various information;
- (3) Less time and financial resources needed to convey results

For better results, this algorithm needs to be constantly upgraded to make the tool's database more diverse, the classification will be more accuracy.

An indicator that reflects the general sentiment of society needs to be thoughtful and calculated in the following studies.