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

Nguyen The Hung
General Statistics Office of Vietnam

Abstract:
Revision of GDP is a necessary work of national statistical offices to ensure that this indicator not only closely reflects the size and potentiality of the economy but also is the basis for calculating many economic and social indicators such as: economic structure, labor productivity, the government's public debt and other relevant criteria. According to UNSTAT, there are three rounds of revaluation of GDP data to deal with inadequacies arising during data compilation including Round 1: Revaluation of quarterly data, Round 2: Reassessment of annual data and Round 3: Review the data periodically.

In which, the third round of major adjustment is usually implemented in certain stages depending on the results of the census, the need for updating the methodology, the need for comparisons as well as the classification tables or according to international recommendations.

In 2019, based on the results of the 2017 economic census and data from the General Department of Taxation on enterprise's information, Vietnam has revised the size of GDP in the 2010-2017 period.

Since the General Statistics Office of Vietnam (GSO) announced the plan to revise the size of GDP, public opinion in the country has had mixed opinions on this issue. Some people doubt the accuracy of the data as well as questioning the independence nature of the statistical office while others support the reevaluation to get a more accurate picture of the economy.

These debates took place in various places, such as seminars, television shows, newspaper articles and social media. In particular, the debate on social network is the most prevalent because anyone can freely express their views on the social network with negligible constraints, the most popular platform is Facebook, thereby causing certain misinformation in the community because the revision of the GDP is one of the in-depth techniques of the national statistics agency. Therefore, listening to public opinion to answer their concerns of social phenomenon is the necessary to make transparent information as well as protecting the reputation of the statistical office.

This study has developed a social listening tool to gather discussions on this matter from fanpage, forum, group, magazines website, personal post, comment, YouTube and then use machine learning and deep learning techniques to analyze sentiments of each by negative / positive and by posting type, relevant poster such as age, level, gender, location to make recommendations to the authorities for more transparent information for the public.

Keywords: deep learning, revision GDP, machine learning, social listening

1. Introduction:

Listening to the people's response about specific issues always plays an important role for strategists and policymakers at the micro and macro levels. At the macro level, listening to people's opinions from any step of the policy formulation helps policymakers make appropriate adjustments with the real situation. When it comes to micro level, business owners need to know the opinion of consumers in order to launch products that match their tastes to better dominate the market.

Normally, listening to people's opinions has been usually conducted through traditional surveys with the limited number of interviewees, the contents and time to conduct, so that the result has certain limitations including the certain fairness (Christopher G. Leggett, 2009).

Nowadays, the great development of science and technology, especially information technology and computer science has brought many ways for people to commutate and express their views on all aspect of the life more easily through social networks, websites, blogs, video, etc. It has created an extra-large amount of digital data known as "Big Data" – "a term that describes the large volume of data- both structured and unstructured - that inundates a business on a day-to-day basis" (the SAS, Inc).

It is more necessary for statistics, academic institutions if they can collect this kind of data, bridging the lagging time of existing statistics and provide an advanced source of data to produce statistics.

The overarching goal of this research is to build up the new database to determine the social response to the revision GDP in the period 2010-2017 in Vietnam from big data source using social listening tool.
To achieve the overall aims, this research seeks out the following specific objectives:

(1) Buiding a social listening tool to collect social views on revision GDP in the period 2010-2017 in Vietnam from the Internet;
(2) Using machine learing, artificial intelligence, deep learning for content analysis;
(3) Determining the social response in term of sentiments, and the source of opinions.

2. Methodology:

According to the report from *Hootsuite, 2019*, with a population of approximately 95 million people (ranked 15th in the world, of which, Internet penetration accounts for more than 60%), Vietnam is the 16th country in the world in terms of Internet users. The internet has become a routine activity with average 7 hours per day participating in Internet-related activities. Along with that, the number of social network users in Vietnam has also increased dramatically. In 2015, the actual number of Facebook users (active social users) recorded by *We are social* was 35 million and to 55 million by 2017. However, in terms of the total number of discussions, within 2015-2017, the number of discussions increased from 20 billion to 55 billion, more than 3 times regarding to the degree of increase. It creates a huge amount of data for researching.

However, information from social networks cannot be collected and searched by common searching engines like Google. Therefore, social listening tool is needed to build for collecting information from both website and social network. The design of tool flows as below:
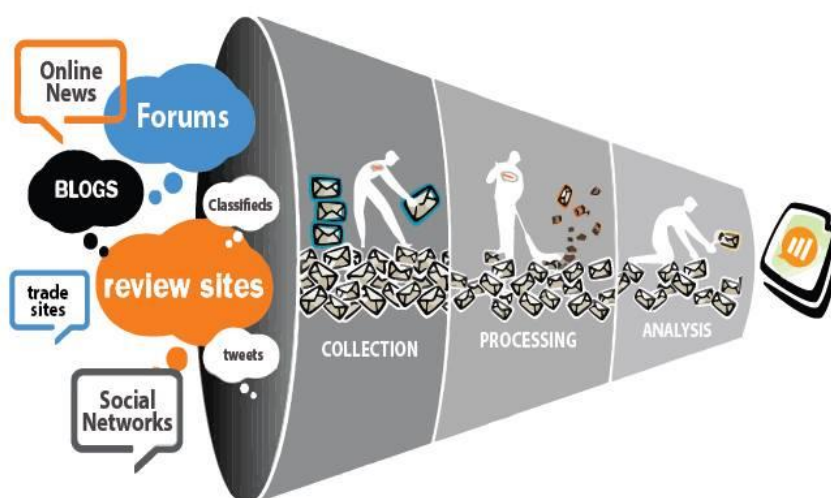

Figure 1: Social listening flow

This tool will collect data from: (1) website; (2) personal post; (3) fanpage post; (4) group post; (5) comments; (6) forum and (7) YouTube. It is larger source than Google with searching only on website.

In this study, the range time from 2019, 16st August (the first time the General Statistics Office of Vietnam - GSO mentioned about revising GDP) to 2020, 31th January (01 month after GSO published the result) was focused on to collect all mentions regarding to the revising GDP of Vietnam in period 2010-2017.

After collected data, the algorithms of machine learning and deep learning will analyze the content of the discussion in each context to determine the sentiment according to three criteria:

- 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.

3. Result:
From the collected database, it is found that there are 1.253 mentions about the revising GDP.
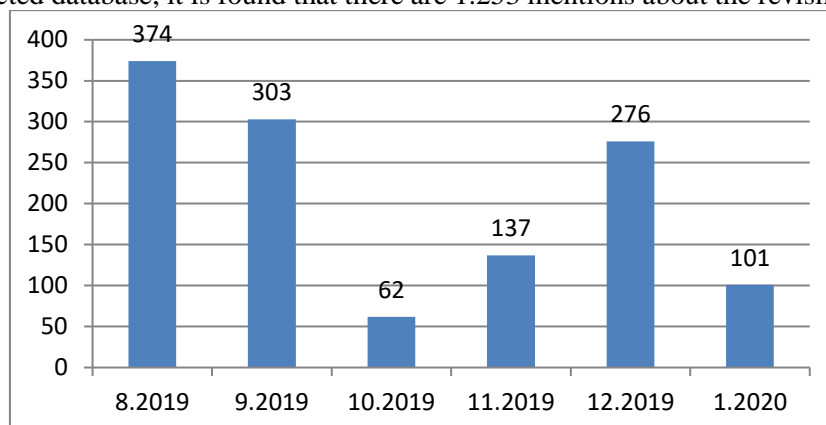

Figure 2: The number of mention by month

The number of mentions varies from month to month, with people starting to pay highest attention to revising GDP in December 2018, when the GSO starts announcing plan to revise GDP in period 2010-2017. The public still watched out this issue in September, 2019 with the minor less mentions than August. It downed after that and rose up again when GSO published the result of this plan.
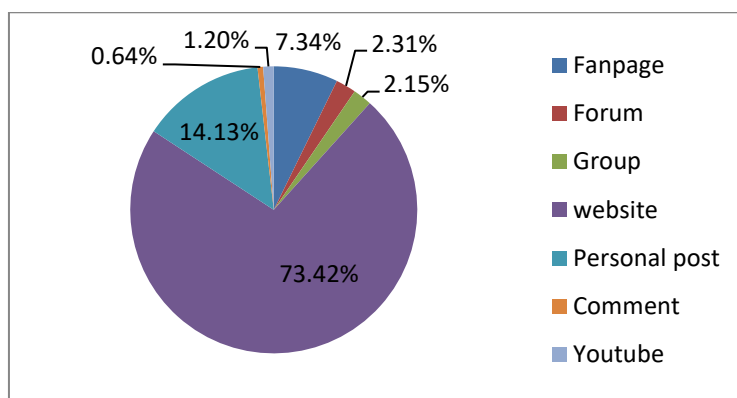

Figure 3: Sources of mentions

When it comes to the sources of mentions, there were noticeably higher from the website at 73,42% and personal post at 14,13%, followed by fan page at 7,34%.

In order to deeply understanding each mention, the algorithms of machine learning and deep learning is used to analyses the sentiment. The sentiment of each mention expresses the society's opinion on revising GDP though criteria: negative and positive (the neutral could be considered as noise).

The number of positive/ negative sentiments also varies from month to month. Contrary opinions appear most often in the early stages when the plan is announced and decreased over time. This proves that most people in society understand this is necessary task and support the re-evaluation of the size of GDP.

At overall, the positive sentiment is always higher than negative as the number 87,21% and 12,79%.
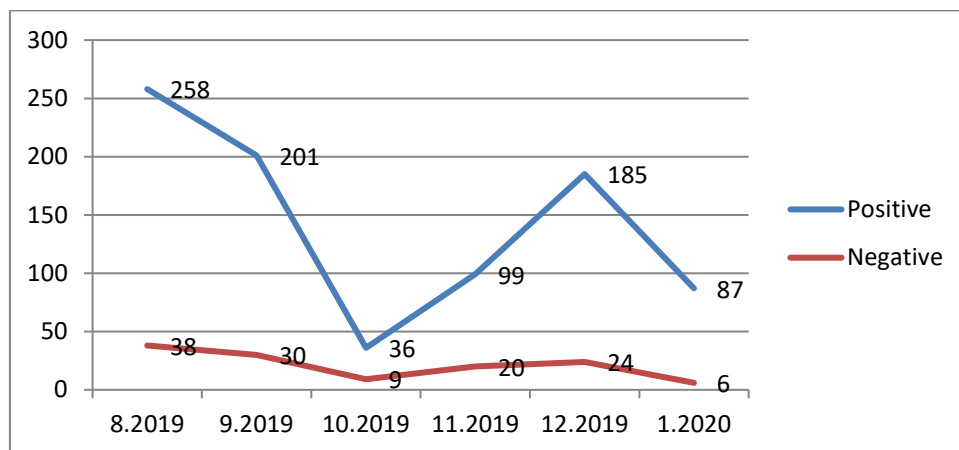

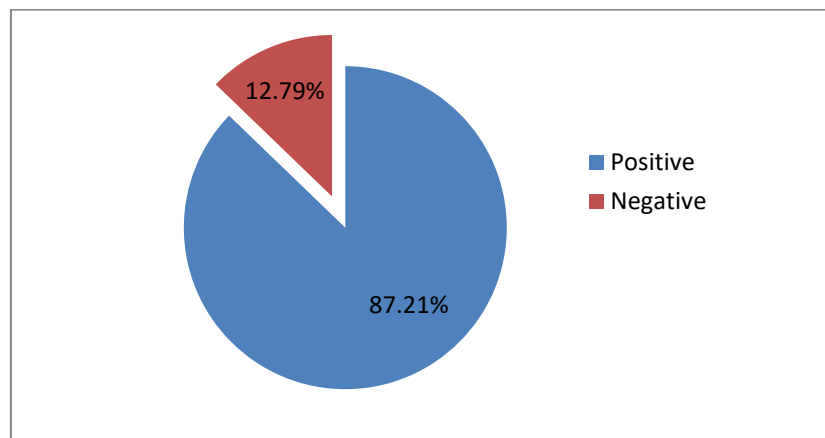Figure 4: The proportion of sentiments from December, 2018 to September, 2019

Figure 5: The proportion overall of sentiment

Based on the personal information of each social network accounts like gender, education, age, location, some distributions of subscriber are illustrated as figure below:
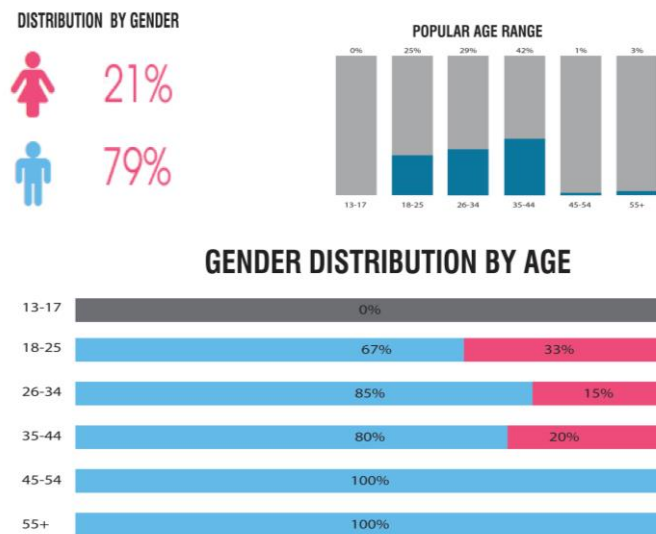
Figure 6: Gender distribution of subscribers

It is clearly that at all ages, men are more interested in macro issues than women (79% vs 21%) and people aged from 26 to 44 are more interested in this issue.
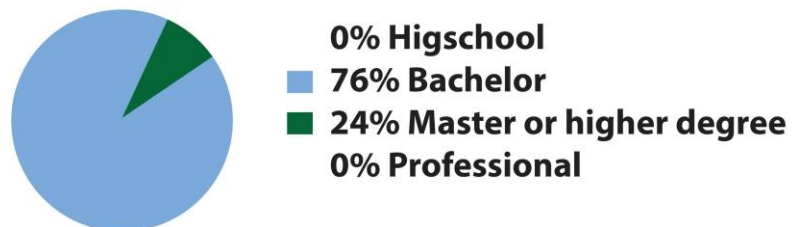

Figure 7: Location and education distribution of subscribers

The urban people are more interested in macro issues than the rural one, which can be explained by the fact that the education level at urban is higher than at the rural.

In addition, re-assessing the size of GDP is an intensive technique so it only attracts people with certain qualifications (from university graduate or above).

4. Discussion, Conclusion and Recommendations:
Revising the size of GDP is a regular task of national statistical offices when they have sufficient information through censuses and other administrative data. However, this reassessment can cause public misunderstandings about the objectivity of the statistical data.

To avoid this situation, the propaganda to public opinion is necessary to minimize negative options that cause instability in the society, reduce trust in statistical products. This propaganda needs to be carried out through official communication channels such as GSO website, mass media and special need to focus on social networks because of its users and information dissemination.

In addition, 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.

From this case it is clearly that social listening is an effective tool for listening to people's opinions about specific issues. This tool provides more diverse information related to research issues than traditional surveys at much lower costs. The database retrieved is the basement for the making policy and decision of authority bodies. This tool can be applied to different levels of management from micro to macro levels depending on the purpose of the researcher.

The application of new algorithms of computer science such as machine learning, deep learning has helped a lot in classifying the sentiment of each mention. However, for better results, this algorithm needs to be constantly upgraded to make the tool's database more diverse, the classification will be more accuracy.

The sentiment of each mention has been separated by month separately. In order to have more tools in management, an indicator that reflects the general sentiment of society needs to be thoughtful and calculated in the following studies.

References:

1. Aqdas Malik, Yisheng Li, Habib Karbasian, Juho Hamari, Aditya Johri (2019): *Live, Love, Juul: User and Content Analysis of Twitter Posts about Juul*, American journal of health behavior.

2. Heather Cole-Lewis; Jillian Pugatch; Amy Sanders; Arun Varghese; Susana Posada; Christopher Yun; Mary Schwarz; Erik Augustson (2015): *Social Listening: A Content Analysis of E-Cigarette Discussions on Twitter*, JOURNAL OF MEDICAL INTERNET RESEARCH 2015 | vol. 17 | iss. 10 | e243 | p. 1

3. Christopher G. Leggett, Naomi S. Kleckner, Kevin J. Boyle, John W. Dufield, and Robert Cameron Mitchell (2009): *Social Desirability Bias in Contingent Valuation Surveys Administered Through In-Person Interviews*, Land Economics.

4. Hootsuite (2019), *Vietnam Digital Marketing report,*

5. Nguyen The Hung (2019): *The introduction the social listening, a new approach for social study*, the presentation at Toulouse School of Economics

6. Tran Thi Kim Thu, Nguyen The Hung (2019: *Using big data retrieved from social listening tool to determine the social response: A case study of increasing electricity prices in Vietnam*, The 2nd International Conference on Contemporary Issues in Economics, Management and Business.