

## **Social Media and Change of Polling**

### **——Argument about question 1**

**Yuxiang Xu 12715948**

#### **1. Introduction**

In the past years, social media has exploded and changed the way people interact with each other. Softwares like Twitter, Facebook, and Wechat take up our time and become the major place where we communicate with others. Social media is the most important place to express our think and opinion.

Ryan, John Barry (2010) found that the key contextual determinant of an individual's vote is the partisan composition of his or her discussion network.

#### **2. Challenges**

Now we are facing the question to predict polling results. What is his support? Will anyone change his support? What will affect people's polling? These questions all need our exploration. In our problem, many challenges are waiting for us. The main challenges include three aspects, data acquisition, data processing, and model building.

The first challenge we have to face is data acquisition. It also can be called data collection. Data is the source and basis of our analysis and modeling. The importance of data for the model is just like the fuel for the engine. If we can't collect enough data, we can't go on our analysis road. Analysis and model have some requirements in data collection. The principal requirement is that data must be big enough. The table we create must include enough rows and enough columns. Enough rows mean more records or more voters. And enough columns means more dimensions about one voter. At the same time, we are facing the requirement of the accuracy of data. We must ensure our data is accurate. Any wrong information will pollute the whole data and lead us to the wrong conclusion. Another requirement is about time. Social media makes the news and messages transfer fast. People's attitude changes fast. We can't establish our prediction in the data for last year.

The second challenge is data processing. After data collection, we have owned much data from many sources. However, the data coming from different sources have a distinct difference. The data may be stored in different formats. The data stored in databases is standard and easy to process. But the data stored in HTML or txt file may be formless. In this step, we need to merge data of different sources into one table. The key to merging is looking for common variables, like phone number, email address, and social security number. The records having the same phone number can be matched together. Contrary to merge, we have to consider filtering. Some information may be irrelevant to our topic. It can

be collected with valid information together. In this step, we need to filter irrelevant data. Totally, we need to clean our raw data and make it easy to model.

The third challenge we have to face is the building model. We collected data, and we cleaned the data. By now, we have some accurate, important, and ordered data. It is a great achievement for us. But it is not enough to come to any conclusion directly. Models can help us to realize our data. Models can help us to predict the user's support.

Except for technical challenges, in spite of the popularity of social media, there are also lots of users who are unwilling to participate in polling, so the results of the survey may be quite different from the real thoughts of users. It is because of the rights for speaking freedoms on the Internet. Thus, some bias in polling may result from some extreme opinions. In addition, The opacity of social media voting may cause the results to be manipulated, just to create topics. This can also make predictions less accurate. These are the external challenges of this problem.

### 3. System

Based on challenges, we have a complete design of our system. The system is consists of three main parts, collection part, processing part, and model part. The collection part is the input of raw data. And the model part has the system output. The processing part is in the middle link. I make a simple map of our system, displaying the main steps.

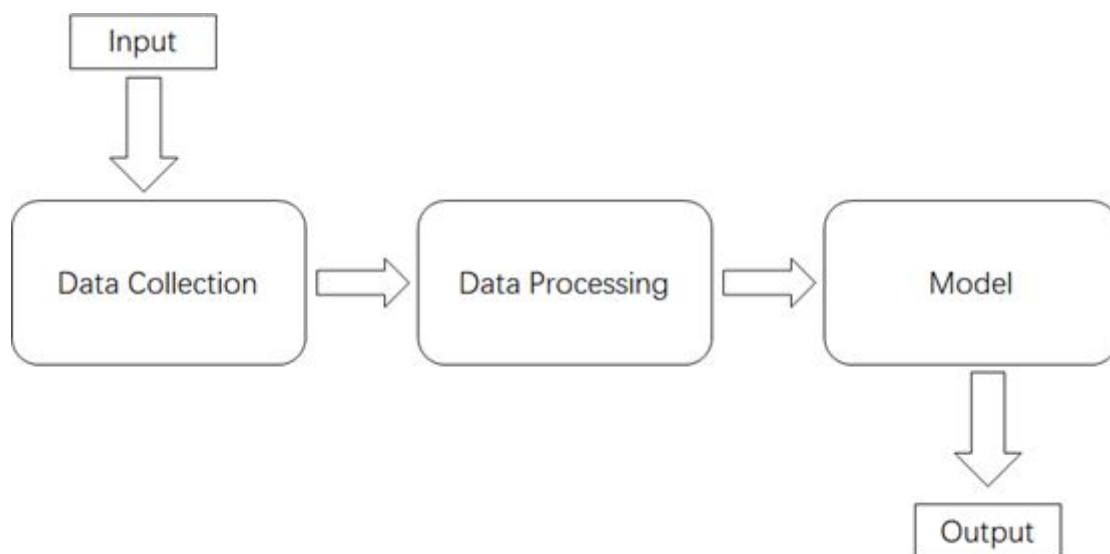


Figure 1: Design of system

As you can see on the map, the collection part accepts data of any source. We can use some tools to help us collect data, such as API, web crawler, and an internet survey. API is the best choice for us. But it needs the support of social media companies. The web crawler is the common method to collect data today. But it is banned by the government and

company sometimes. An Internet survey is also a common approach to collect our data. The disadvantage of an internet survey is that the survey result is subjective. Data processing depends on ETL tools. ETL is short for extract, transform, and load of data. The classical ETL tools include Informatica, Infosphere Information Server, and Oracle Data Integrator. In the model part, we need tools to help to establish a model. Python and R are good choices for us in this step. The two software are open source and useful. It can help us to reduce costs and establish stable models. On the basis of our system, if people express opinions on social media, we can predict the change of their supports.

#### **4. The ethical and social consequence**

The biggest ethical problem is the conflict between our study and people's right of privacy. To make our study accurate, we want to collect more data. The behavior may violate people's right of privacy. The classical example is Facebook. Mark Elliot Zuckerberg was inquired in Congress on Apr 11, 2018. According to privacy, we must collect information as little as possible. In our study, we must establish a clear data boundary. All our data should be under the limitation of the national jurisdiction and permission of the user.

#### **5. Reference**

Gonzalez-Ocantos, E., De Jonge, C. K., Meléndez, C., Osorio, J., & Nickerson, D. W. (2012). Vote buying and social desirability bias: Experimental evidence from Nicaragua. *American Journal of Political Science*, 56(1), 202-217.

Metaxas, P. T., Mustafaraj, E., & Gayo-Avello, D. (2011, October). How (not) to predict elections. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing* (pp. 165-171). IEEE.

Ryan, J.B. The effects of network expertise and biases on vote choice. *Political Communication* 27.1 (2010): 44-58.

Ryan, J. B. (2011). Social networks as a shortcut to correct voting. *American Journal of Political Science*, 55(4), 753-766.

Sokhey, A. E., & McClurg, S. D. (2012). Social networks and correct voting. *The Journal of Politics*, 74(3), 751-764.