*Article*

# Measuring Online Public Opinion for Decision Making: Application of Deep Learning on Political Context

**Daesik Kim** [1,†] , **Chung Joo Chung** [2,†] **and Kihong Eom** [1,*]

1   Department of Political Science and Diplomacy, Kyungpook National University, Daegu 41566, Korea; daesik.kim@knu.ac.kr
2   Department of Media and Communication, Kyungpook National University, Daegu 41566, Korea; cjchung@knu.ac.kr
*   Correspondence: kheom@knu.ac.kr
†   These authors contributed equally to this work.

**Abstract:** Thoughts travel faster and farther through cyberspace where people interact with one another regardless of limitations in language, space, and time. Is a poll sufficient to measure people's opinions in this era of hyperconnectivity? This study introduces a deep learning method to measure online public opinion. By analyzing Korean texts from Twitter, this study generates time-series data on online sentiment toward the South Korean president, comparing it to traditional presidential approval to demonstrate the independence of the masses' online discourse. The study tests different algorithms and deploys the model with high accuracy and advancement. The analysis suggests that online public opinion represents a unique population as opposed to offline surveys. The study model examines Korean texts generated by online users and automatically predicts their sentiments, which translate into group attitudes by aggregation. The research method can extend to other studies, including those on environmental and cultural issues, which have greater online presence. This provides opportunities to examine the influences of social phenomenon, benefiting individuals seeking to understand people in an online context. Moreover, it helps scholars in analyzing those public opinions—online or offline—that are more important in their decision making to assess the practicality of the methods.

**Keywords:** online public opinion; offline public opinion; polling; deep learning; BERT

## 1. Introduction

People's preferences form the public's collective sentiment and various political elements including elections, representation, and policymaking. Public opinion is a group expression or consensus of people who share the same or similar interests [1] (MacDougall, 1952). Naturally, distinguishing interest groups in the realm of politics is challenging. People can have more than one preference and be part of multiple interest groups simultaneously. Until today, polling has been the dominant method to assess public opinion. Specifically, presidential approval is a good example of a polling estimate to measure public opinion in politics.

Presidential approval is widely used among different countries to explain how much public support an incumbent leader of a state commands. This popular measure has gained importance and influence since Gallup asked the question, "Do you approve or disapprove of the way the incumbent is handling his job as president?" in the 1930s [2]. It has become one of the most essential indicators that explain the state of political affairs. News outlets emphasize the ups and downs of presidential approval and discuss different reasons for changes in the ratings. Meanwhile, the public pays attention to this performance measure. Thus, presidential approval influences how people perceive the current state of politics.

Numerous pollsters may have significantly different results depending on their detailed polling methods. The proportion of cellular phones in a sample is controversial, as it

can alter polling results in a particular direction. Lack of response also raises doubts about a poll's representativeness. For example, in 2016, Donald Trump won the U.S. presidential election despite all election surveys predicting a victory for Hillary Clinton. In South Korea's general elections in 2020, the ruling party achieved a landslide victory by winning 180 out of 300 seats; similarly, no poll predictions came close to the actual outcome. Therefore, the public's doubt toward election polls and forecasts has increased since the election in Korea. Understanding the aggregated will of the public is difficult and becomes even more challenging with the rapid lifestyle changes in the era of hyperconnectivity.

Methods of gauging public opinion have not changed much in the real world, yet the popularity of smartphones has changed how people live. Specifically, social network services have influenced people's communication behavior. As people increase their use of mobile chats and social network services to communicate, responses to traditional voice platforms, such as telephone surveys, decrease. Considering changes in communication, it is critical to respond to the hyperconnected environment in the future by acknowledging public opinion of the mass in cyberspace. The Internet's pervasiveness in our everyday lives affects politics. More politicians have been using online channels to engage with the public. Former presidents Trump and Barack Obama have been using Twitter, and South Korean politicians have mainly used YouTube for their political communication. The Internet has become an essential element of not only political communication but also of election campaigns. Obama's and Mitt Romney's presidential campaigns actively utilized the Internet and social media in the 2012 U.S. presidential elections [3]. This trend has continued to the recent presidential elections in the United States, including the Trump campaigns in 2016 and 2020 [4]. Modern election campaigns vigorously appeal to their supporters through various online services. Specifically, social networks have gained significance as a medium to facilitate political movements. Online social outlets have aided many public protests around the world, for instance, the Arabs' prodemocracy movement and the civil rights movement in the United States [5,6]. Although the Internet has deeply penetrated the realm of politics, current measures of presidential approval do not include online platforms.

This study aims to gauge online public opinion using textual data from the Internet, specifically Twitter. It introduces a deep learning technique to measure online political public opinion and explains whether the online public is independent from the offline one. This adds peculiar values to existing studies. First, the applied deep learning technique demonstrates a method to extract sentiments from user-generated texts. This is particularly important for languages such as Korean, wherein words change their meanings and grammatical features based on their synthesis form. Second, the study extends the application of the method to politics, specifically the evaluation of a government. Online public opinion can provide complementary insight into the conventional offline presidential approval. Third, measuring public opinion on various issues using online data requires much fewer resources in terms of time, labor, and capital. This indicates a wealth of data available in the future for research on public opinion by reducing the temporal and spatial limitations of an offline political survey. Overall, the study involves explorative research regarding the application of deep learning to political public opinion in an online environment to facilitate more effective decision making.

## 2. Literature Review

### 2.1. Measuring Public Opinion in Politics

Public opinion refers to the ideas, thoughts, expressions, interests, or beliefs of particular people who are part of broader society [1,7,8]. The researchers aim to understand what people think. Polls have made it possible to represent the public's aggregated attitude and have added value to politics by providing technical and organized information to the public, politicians, and researchers [1,9,10].

Polls measure the public opinion of a target population. In his work, MacDougall (1952) clearly explained the boundaries of public opinion. Geographical distinctions define

the scope of a public, which means multiple publics can exist in the world. Recent years have seen more spatial separations as many different services have become available in cyberspace. A person can have various interests and participate in different interest groups, which, according to MacDougall [1], are equal to many participating publics. Difference in thoughts is another reason why the public is not unique. Polling or surveys are the dominant methods to gauge public opinion of the mass despite their shortcomings [11,12]. Berinsky [11] emphasizes that political scientists must be cautious about their choice of sample and the questions to be asked, indicating the difficulties in creating a suitable sample for a target population and extracting meaningful results through a proper question. Koo [12] specifies the difficulty of having a representative sample in a political environment in South Korea. His study demonstrates that young female voters are under-represented in samples for an election prediction. Both authors highlight the transformative influence of mobile phones and the Internet on people's lifestyles as a reason for inaccurate samples.

Presidential approval has been the most popular aspect when gauging public opinion in politics. Researchers have studied the subject since John Mueller's seminal study in 1970. Reviews of presidential approval fall under two main branches: effect and cause. An example of the former would be the influence of presidential approval on the president's policy proposals [13], public positioning [14], presence in the legislative body [15], and legislation success. The latter branch, meanwhile, considers presidential approval as a dependent variable and examines what influences public opinion. As Mueller wrote in his book titled *War, Presidents, and Public Opinion* [16], for example, war is a driving factor for presidential approval. Prolonged war and a high death count, especially among the U.S. military, cause a decline in approval. This factor was confirmed in other studies (Gartner & Segura, 1988; Ostrom & Simon, 1985). In addition, economic conditions significantly affect presidential approval, as many studies have found [17–20].

### 2.2. Online Public Opinion and Its Methods

This study addresses the problem of an offline survey by measuring the mass opinion available in cyberspace. It explores a way to extract group sentiments using user-generated texts and a deep learning technique. Online public opinion literature has two branches. The first is a group of studies investigating distinctive characteristics of online public opinions. This category explains the extent to which the Internet represents the public. Duggan and Brenner [21] reveal that social network platforms have different user compositions, which leads to a distinctive level of general population representativeness. Moreover, this trait is not specific to the online environment in the United States. Mellon and Prosser [22] argue that British users of Twitter and Facebook share no similarities with the general population; they differ in many factors including age, gender, and education level. Some studies argue that in South Korea, social networks represent a particular group of people rather than the general population [23,24]. In addition, scholars have attempted to analyze the political traits of Twitter users. Cyberspace users demonstrate strong political engagement and partisanship [25,26]. Online services can underrepresent specific groups such as women, as well as certain political ideologies [26,27].

The other branch seeks to interpret political phenomena using online data. The most considerable interest is in predicting election results using social media [28–30]. Related studies analyze different signals to calculate the possibility of election result predictions. Another area of interest in politics is the subject of issue saliency. Similar to presidential approval literature, these studies illustrate particular themes influencing election results: election debates [31,32] and economic status [33]. They all explain that these factors can shape elections and presidential approval.

The above studies represent research interest in social media and its influence, yet they do not completely understand the online public's thoughts. The lack of research here is due to the difficulty in collecting and processing massive volumes of unstructured data. If it were possible to utilize such data, information from the Internet can be a great complement

to existing measures of public opinion. There is a constant real-time inflow of information, as people continuously communicate in the online environment.

Online data is fundamentally different from traditional survey data. The former does not follow the existing structure, which consists of a question and an answer [4]. Unlike in a survey, useful information is scattered and hidden under big data, which refer to both a vast amount of data and a multivalent process facilitating the combination of heterogeneous data and the extraction of valuable information for use [34]. Therefore, techniques for handling big data should be different from the ones used in traditional research. There are mainly two approaches to extract the aggregated attitudes of people who use online data: the counting method and sentiment analysis [4]. The first method involves the simple counting of texts with a particular pattern yielding mixed results. Some studies have illustrated the successful prediction of elections [30,35], while others have explained that counting does not reveal much predictive power [36].

The other approach is sentiment analysis, which aims to understand emotion hidden in a text through the use of a computer. The analysis tool takes raw text data, tokenizes texts, and analyzes the processed words [37]. There are supervised and unsupervised learning methods available for sentiment analysis. The supervised method uses training data, which contain predetermined emotions regardless of a subject domain, and eventually builds a model predicting uncategorized text data. Neural networks introduce substantial improvements in natural language processing, which leads to better sentiment classification. Bidirectional Encoder Representations from Transformers (BERT), a pre-trained neural network, exhibits considerably higher performance than other sentiment classification tools [38]. The unsupervised method, meanwhile, utilizes already established lexicon or dictionary and sentiment categories. Many studies on online communication have incorporated unsupervised learning methods [38–42]. Table 1 summarizes the aforementioned methods, which support extracting sentiment from online texts.

**Table 1.** Summary of sentiment analysis methods.

| Type | Strength | Weakness |
|---|---|---|
| Counting Method | - Simplicity<br>- Easy to apply<br>- No advanced technology required | - Low and unstable accuracy |
| Sentiment Analysis Unsupervised Learning | - High and stable accuracy with a proper dictionary<br>- No advanced technology required | - Domain-specific approach<br>- Difficulty in building a lexicon dictionary<br>- Performance reliant on the quality of a dictionary |
| Sentiment Analysis: Supervised Learning | - Most recent and advanced<br>- Universality of a model<br>- Less data preprocessing | - Advanced technology required<br>- Difficulty in building relevant training data |

This study performs sentiment analysis on Korean texts collected from Twitter through a trained neural network. It focuses on presidential approval, the most popular measure of public opinion in the political context, and highlights South Korea. The analysis aims to answer two main questions:

Q1. How do we collect and process an extensive amount of unstructured data and user-generated non-English texts to measure aggregated sentiment?

Q2. Does the measured online public opinion represent the population of a survey in political context?

## 3. Methods

### 3.1. Data

This study collects data from Twitter, a popular source for academic research because it has sufficient users worldwide and researchers can access good-quality raw data from it. Unlike other social services such as Facebook, it has an official gateway to retrieve users' text with a greater amount of subsidiary information including language, location, and related texts. There is a limit of 280 characters to how much a user can write in a single post; therefore, sentences are naturally a base unit for data translation. According to a report from the Korea Information Society Development Institute [43], 14% of all SNS users actively engaged on Twitter in 2018, of whom 12.4% were women and 15.5% were men. Gender distribution on Twitter is relatively balanced compared to other SNS, such as Facebook and Kakao Story This study uses collected data in two main parts: neural network training and sentiment prediction. The only difference between these two processes is that the former requires sentiment labels by human coders.

This study collects real-time livestreamed tweets in the Korean language filtered by the keyword *Moon Jae In*, using Twitter's application protocol interface (API). (The study uses API version 1.0 for data collection. Twitter launched API version 2.0 in November 2021, which allows researchers full access to its archive. The study uses the first version of the API for data collection). A computer continuously sends maximum requests to the Twitter server every 15 min, and the server dispatches randomly aggregated batches of tweets upon the request. This process generated a total of 7,253,878 tweets for 2019. This dataset has two distinguishing qualities: one is that the collected texts are limited to 140 characters, and Twitter removes characters after the maximum length when it dispatches the data. The other is that it contains many replied tweets; in this case, 628,040 texts are retweets, a relayed text, comprising 7.25% of the entire dataset.

Neural network training needs data with relevant labels. The machine undergoes supervised learning using data in a "text–sentiment" format. The trained network attempts to replicate the classification similar to the training data, which are a subset consisting of 10,000 tweets in the entire dataset. Five coders who have the same set of texts place sentiment labels individually depending on the superficial interpretation of a text. A coder decides on a sentiment from three categories: positive, negative, and neutral. The final sentiment of a sentence becomes the mode of all five coders. If a text has multiple modes, the final prediction considers the following tie-breaking rules: a text has a neutral sentiment if two modes are bipolar. Any combination with the neutral indicates a direction among the coded sentiments. This direction becomes the final label of the sentence. For example, if there are two neutral and two positive coders, then the direction becomes positive, and the final sentiment is also positive. Among the 10,000 coded tweets, 21.55% have a unanimous opinion, and 85.58% have only one mode. Texts with bipolar modes are 1.53% of the total training data. The training dataset has approximately twice more negative tweets (43.76%) compared with the neutral and positive ones. The percentages of neutral and positive texts are almost identical at 28.68% and 27.56%, respectively.

### 3.2. Methods

Unsupervised and supervised models can perform sentiment classification as explained in the literature review. This study tests supervised machine learning models including convolutional neural network (CNN), recurrent neural network (RNN), and BERT, as they perform relatively better than the unsupervised model and the traditional logistic regression-based supervised model [38]. The deep learning approach suits the Korean language better than lexicon-based unsupervised models for several reasons. First, Korean does not have a well-defined lexicon dictionary for sentiment analysis. The study attempts to apply sentiment analysis to Korean politics, which indicates that lexicon data for the unsupervised model should fit political context in South Korea. Second, morphological analysis is difficult for Korean as it is an agglutinative language (wherein a word can change its meaning depending on its neighboring affix).

The study considers embedding type, embedding size, and a neural network to perform supervised learning, specifically deep learning. Embedding is a process for converting words to vectors, which a computer can understand. This conversion is performed in different ways; the study uses Word2Vec and FastText. Embedding size refers to the dimensionality of embedding vectors and is a factor associated with the resolution of natural language complexity. This study uses embedding sizes of 100, 200, and 300 (BERT is a pre-trained model that includes its own unique embedding type and dimension). A neural network is a supervised deep learning algorithm that serves as a simple replication of the human brain. The study tests three main neural networks: CNN, RNN, and BERT [44–48]. (The study applies the gated recurrent unit (GRU) through RNN and implements modified BERT by adding linear layers at the end. For the detailed explanation of all neural networks used in the test, LeCun [47] explains CNN, Cho et al. [44] introduce GRU algorithm of RNN, and Vaswani et al. [48] illustrate BERT). The BERT model is a pre-trained algorithm, and this study uses KoBERT, which is pre-trained through Korean texts [49]. (The KoBERT GitHub page [49] contains the parameter information about the model and its code. The study applies transfer learning process to KoBERT model to perform a political sentiment classification on Korean Twitter texts). Finally, this study tests both two- and three-category classification; this means that the network distinguishes either between positive and negative or between positive, neutral, and negative. When the network is trained to perform two-category classification, unclassified tweets become neutral ones. In summary, there are three embeddings with three-dimension sizes and three neural networks with two classification options. Combinations of these factors can yield different accuracy level of the trained model.

Herein, two embedding types, three embedding dimensions, three neural networks, and two groups of classification category lead to the total of 26 combinations to test. (The study uses Python to pre-process Twitter data and PyTorch to construct neural networks. The embedding process uses the Gensim package through Python). The test result applies the most appropriate logic to the machine and obtains daily sentiments from all tweets collected in 2019. Table 2 includes accuracy figures for all parameter combinations.

Supervised learning requires three datasets: training, validation, and test. This study divides the entire dataset into three ratios: 75%, 12.5%, and 12.5%. The most substantial portion is for the training of the network. The other sets are for validation and testing, which measure the system's feasibility during and at the end of the training. The validation process occurs at points during training using a pre-assigned portion of data to check whether an algorithm is properly learning from data. The test set is used to examine the final performance of a trained model; therefore, it remains untouched until the completion of algorithm training. The accuracy score, which refers to the percentage of correctly predicted data, is from the test set and determines the performance of the analysis method.

According to the accuracy scores in Table 2, the networks accomplish better results in the two-category classification task; that is, reducing a classification category improves accuracy in all combinations. The study sets the threshold probability to 0.7 to determine whether a tweet reveals a sentiment. The BERT model yields the best result in the three-category task with 84.92% accuracy on the test set. For two-category classification, the best network is RNN with Word2Vec, 300-dimension, at 94.26%. The BERT model exhibits 94.18% accuracy, which is 0.08% lower than the RNN model (Other performance measurements of the applied BERT model including precision, recall, and F1 score are 0.919, 0.924, and 0.922, respectively. The hyperparameter settings for the deployed BERT model are 12 layers, 768 hidden layers, and 12 self-attention heads.). A larger embedding dimension size does not guarantee better performance. For example, CNN with an embedding size of 200 tends to have higher accuracy except for the combination of RNN, FastText, and 100-dimension in the two-category classification task. Between Word2Vec and FastTrack, it is impossible to conclude whether a particular embedding is better for this sentiment classification project. Word2Vec is a better match for CNN, while FastTrack generally yields

better results with RNN. Overall, there is no outstanding branch in terms of accuracy, which varies depending on different component mixes.

**Table 2.** Accuracy of all tested models.

| Order | Classification | Embedding Type | Embedding Dimension | Neural Network Model | Accuracy |
|---|---|---|---|---|---|
| 1 | 3 | Word2Vec | 100 | CNN | 82.64 |
| 2 | 3 | Word2Vec | 200 | CNN | 84.40 |
| 3 | 3 | Word2Vec | 300 | CNN | 84.08 |
| 4 | 3 | Word2Vec | 100 | RNN | 82.48 |
| 5 | 3 | Word2Vec | 200 | RNN | 80.56 |
| 6 | 3 | Word2Vec | 300 | RNN | 77.20 |
| 7 | 3 | FastText | 100 | CNN | 78.96 |
| 8 | 3 | FastText | 200 | CNN | 80.56 |
| 9 | 3 | FastText | 300 | CNN | 80.40 |
| 10 | 3 | FastText | 100 | RNN | 81.92 |
| 11 | 3 | FastText | 200 | RNN | 79.84 |
| 12 | 3 | FastText | 300 | RNN | 83.04 |
| 13 | 3 | BERT | - | BERT | 84.92 |
| 14 | 2 | Word2Vec | 100 | CNN | 91.94 |
| 15 | 2 | Word2Vec | 200 | CNN | 93.38 |
| 16 | 2 | Word2Vec | 300 | CNN | 92.84 |
| 17 | 2 | Word2Vec | 100 | RNN | 91.72 |
| 18 | 2 | Word2Vec | 200 | RNN | 91.72 |
| 19 | 2 | Word2Vec | 300 | RNN | 94.26 |
| 10 | 2 | FastText | 100 | CNN | 90.95 |
| 21 | 2 | FastText | 200 | CNN | 90.62 |
| 22 | 2 | FastText | 300 | CNN | 91.50 |
| 23 | 2 | FastText | 100 | RNN | 92.16 |
| 24 | 2 | FastText | 200 | RNN | 91.50 |
| 25 | 2 | FastText | 300 | RNN | 92.20 |
| 26 | 2 | BERT | - | BERT | 94.18 |

This study analyzes the sentiment of all collected tweets in 2019 using the BERT model customized for two-category classification. Among the tested branches, BERT yields high accuracy scores for all given tasks with relatively stable performance. BERT is the most advanced neural network among the systems tested in the analysis, and it is designed to handle complex sequential data such as natural languages [38]. It is also pre-trained in such a way that it does not require extra components such as embedding. It is the more straightforward system to deploy in this sentiment classification task compared to the others. Overall, the study includes a sequential process to extract online public opinion. First, a machine automatically collects user-generated texts, Twitter for this study, on a particular subject for a given period. Second, human coders determine sentiments of sample texts without cooperation, and majority rule decides a final sentiment label of a text. Third, the labeled data train a deep learning model to build a sentiment classifier. The study tests different algorithms with various factors, including embedding type and size, and deploys the modified KoBERT model to analyze user-generated Korean texts within a political context. Finally, all collected online texts become the online public opinion of a certain subject in aggregation. The study statistically compares the online sentiment and offline public opinion on a similar issue to examine the uniqueness of online public opinion.

## 4. Analysis

The present study utilizes supervised deep learning to extract public opinion from the collected tweets. The trained neural network processes Twitter texts and generates sentiment predictions. As explained in the previous section, the BERT model performs sentiment analysis on user-generated Korean texts. Before entering the network, the data requires cleanup. This preprocessing stage includes removing unnecessary words, punc-

tuations, and special characters that offer no information to help determine user attitude. The modified BERT model calculates the probability of whether a tweet's sentiment leans toward a positive or a negative feeling. Figure 1 presents the time-series graph of daily aggregated sentiment in 2019, which shows 190 days with more negative sentiments and 175 days with greater positive sentiments. This trend reveals an extremely high volatility, making it difficult to acknowledge a potential pattern.
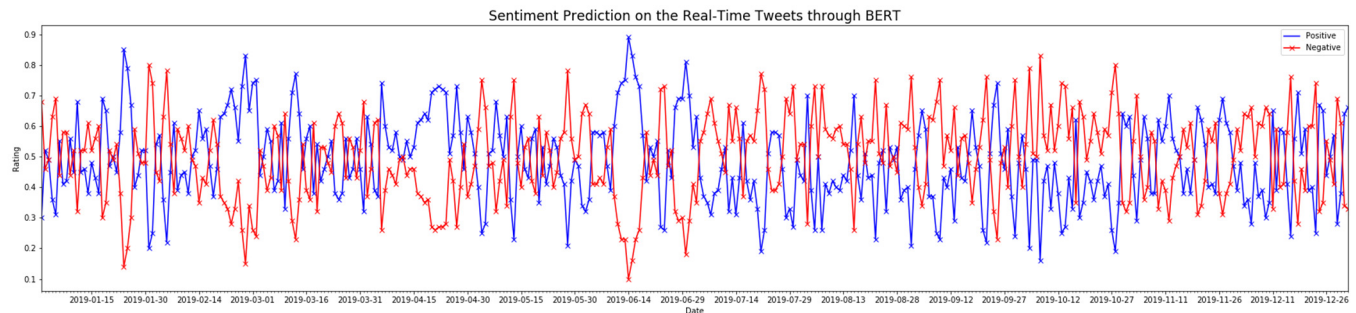


**Figure 1.** Daily aggregated sentiment in 2019.

Figures 2 and 3 show the weekly and monthly transformations of the predicted sentiments, respectively. The graphs reveal that volatility significantly decreases from the daily measurement. While the online public had a positive feeling in the first half of 2019, the general sentiment changed in the second half. There were more positive events in the first half of the year, which included the third summit meeting between South and North Korea on 18 September 2018. This hopeful ambience generated by the consistent engagement between the two nations continued to 2019. Meanwhile, the leaders of the United States and North Korea again had a surprise meeting at the Joint Security Area of Pannumjeom briefly after the G20 meeting in 2019. This series of foreign affair events created a positive sentiment among the online public. In addition, in 2019, the 100th year of Samiljeol, Independence Movement Day, was commemorated, the movie *Parasite* won an Oscar, and one of the South Korean national soccer teams placed second in the U20 World Cup in the first half of that year.
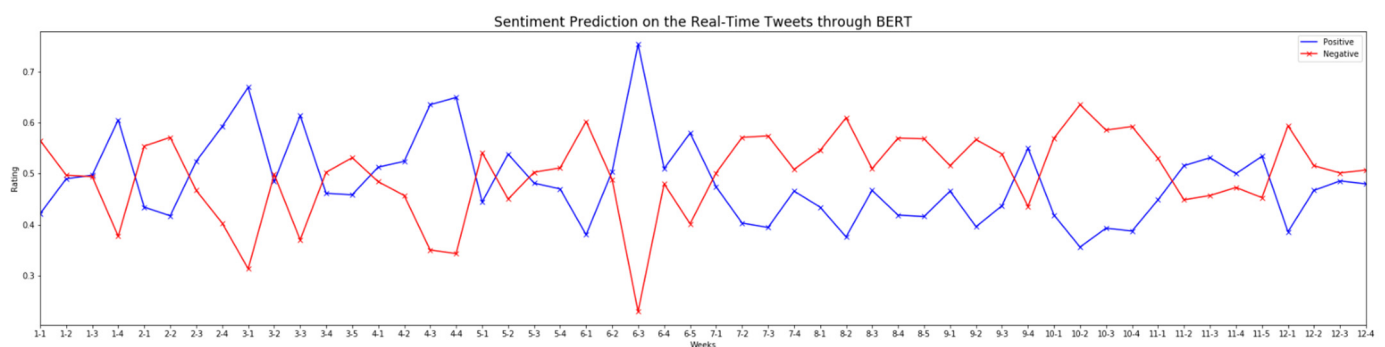


**Figure 2.** Weekly transformation of online sentiment in 2019.

However, the second half of 2019 began with Japan's export regulations on semiconductor materials, which Japanese companies sell to South Korea. In August, South Korean President Moon Jae In appointed university professor Guk Cho as minister of the Justice Department, which caused massive outrage over suspicions that his family might abuse his social authority. In December, the government laid stringent regulations on the real estate market. In addition, the National Assembly passed a law to create an independent investigation agency targeting high-ranked public officers. Indeed, the events in the second half of 2019 are more controversial than the ones in the first half. Specifically, the online public's attitude toward the president was significantly shaped by the bickering over issues

involving the former minister of justice and the installation of an investigative body for top government officers.
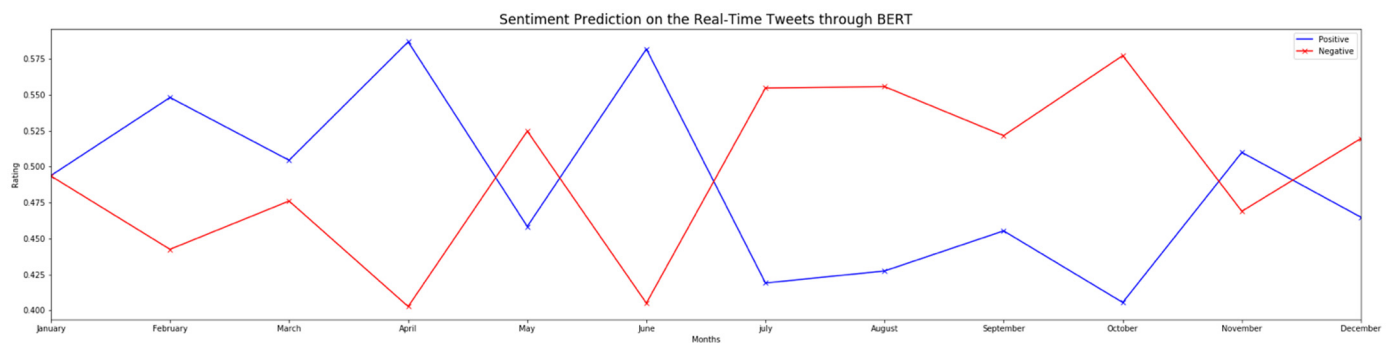


**Figure 3.** Monthly transformation of online sentiment in 2019.

How do these online sentiment trends relate to traditionally measured public opinion? The analysis examines correlations between online and offline measures of presidential approval. Two polling agencies, Gallup and Realmeter, regularly report presidential approval in South Korea to the public. Figure 4 presents the presidential approval ratings measured online and offline. Neither polling agency has everyday values pertaining to presidential approval. Realmeter has sufficient daily polls compared to daily online sentiment values. The graph compares the online sentiment and Realmeter's daily poll. Both illustrated trends show high volatility.
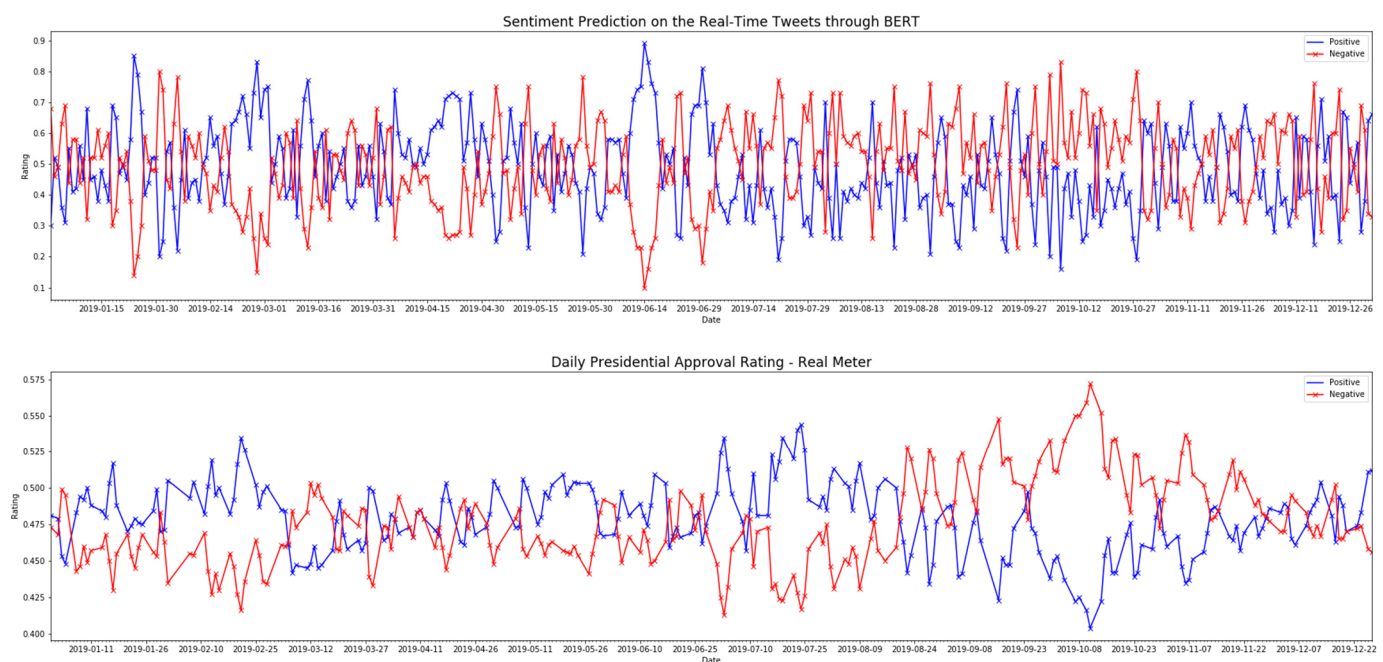


**Figure 4.** Comparison of daily online sentiments and Realmeter's poll.

Table 3 shows correlations between daily online and offline values. Between the online trend and Realmeter's poll, correlation coefficients are 0.16 for negative sentiment and 0.13 for positive sentiment. Although these correlations are close to zero for the whole of 2019, certain periods exhibit a similar pattern, such as the window between January and February, in which public attitudes of both environments display common paths. Positive feeling increases at the end of January and February. Positive sentiment significantly decreases at the beginning of March but substantially increases at the end of June. According to the poll, this rapid increase in positive sentiment also happens online.

**Table 3.** Correlation between online sentiments and Realmeter's daily poll.

| | Online Sentiment | |
|---|---|---|
| | **Positive** | **Negative** |
| Realmeter Daily Poll | 0.127 (0.342) | 0.164 (0.273) |

Standard errors in parentheses.

Online and offline data have fundamental differences. Polls have a finite number of people in a sample, whereas online collections do not restrict the number of tweets a machine can collect. The total number of texts is different for each day, which naturally increases volatility. Polls conduct surveys on multiple days to generate more stable values. For example, Realmeter and Gallup use 3-day average values (Gallup also uses a 2-day average value instead of a 3-day figure when the survey period includes a holiday). To compare sentiment values between online and offline data, the study converts daily online sentiments into weekly ones by averaging the online values on dates when offline polls are available. Figure 5 illustrates the weekly trends in public opinion on the president. Polls from the two agencies have an almost identical graph, but the online trend is different from the offline one.
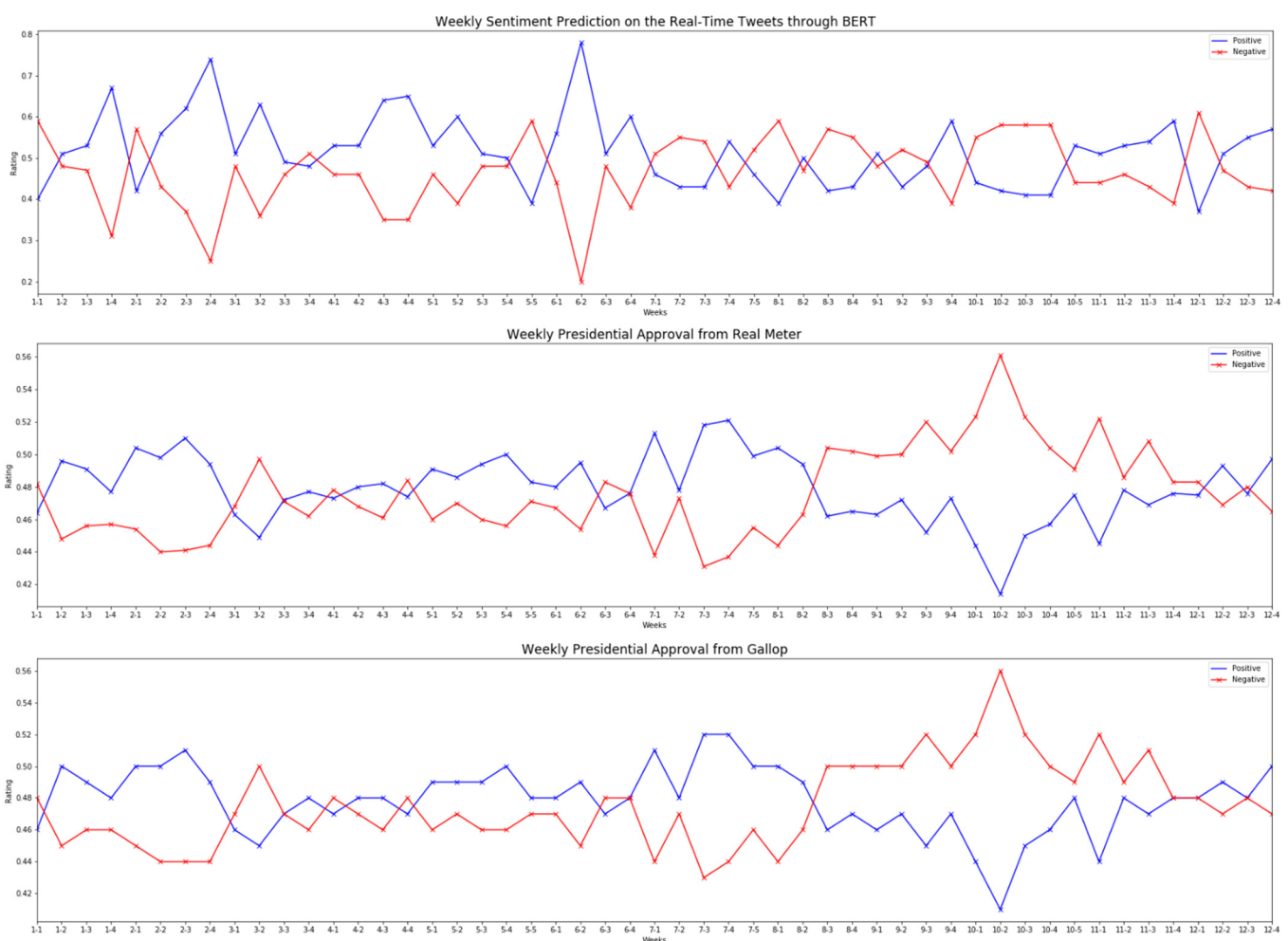


**Figure 5.** Comparing weekly ratings between online and offline.

Weekly sentiments have much less noise compared to the daily graphs. Table 4 shows the correlations among these graphs. For the positive and negative sentiments, the real-time tweets and the Realmeter poll have correlations of 0.19 and 0.28, respectively. The weekly

data has higher coefficients than the daily data, indicating that volatility reduction only improves the correlation with very limited level.

**Table 4.** Correlation between datasets and the Realmeter and Gallup weekly polls.

| | Online Sentiment | |
|---|---|---|
| | **Positive** | **Negative** |
| Realmeter's Weekly Poll | 0.193 (0.607) | 0.278 (0.449) |
| Gallup's Weekly Poll | −0.100 (74.233) | 0.066 (58.458) |

Standard errors in parentheses.

Between online and Gallup data, the coefficients are lower, −0.1 for the positive sentiment and 0.07 for the negative sentiment, and they exhibit no relation.

Polls ask people direct questions on a particular issue. Meanwhile, online data have no voluntary control; a machine simply collects available data from the Internet that fit the scope of a subject. This fundamental difference may lead to a low level of correlation between online and offline sentiments. The analysis examines two more comparisons to understand how these online and offline public opinions are different. Online sentiments can represent offline public opinion with time differences. Simply put, the way people think in the real world can manifest in the online world sooner or later. The study measures the correlations of both online and offline data with time adjustments. Table 5 illustrates the relation between the two types of public opinion with various time differences.

**Table 5.** Correlation between the polls and online data with time adjustments.

| | | Realmeter | | Gallup | |
|---|---|---|---|---|---|
| | | **Positive** | **Negative** | **Positive** | **Negative** |
| Online Sentiments | t − 1 | 0.196 (0.655) | 0.298 (0.477) | 0.304 (0.006) | 0.366 (0.005) |
| | t − 2 | 0.194 (0.679) | 0.307 (0.494) | 0.310 (0.006) | 0.386 (0.005) |
| | t − 3 | 0.154 (0.664) | 0.273 (0.484) | 0.050 (0.007) | 0.110 (0.005) |
| | t + 1 | 0.180 (0.579) | 0.288 (0.420) | 0.040 (0.005) | 0.103 (0.004) |
| | t + 2 | 0.135 (0.593) | 0.236 (0.429) | 0.127 (0.006) | 0.143 (0.006) |
| | t + 3 | 0.026 (0.643) | 0.162 (0.484) | 0.212 (0.006) | 0.254 (0.005) |

Standard errors in parentheses.

Between Realmeter data and online sentiments, most improvements occur 1 and 2 days before the target date, t − 1 and t − 2. The correlation increases by 0.02 and 0.03 points for negative sentiments, respectively, and 0.003 and 0.001 points for positive ones, respectively. The Gallup correlation increases with time adjustments. Two days before the target date, t − 2, reveals the largest difference in correlations. The Realmeter and Gallup results illustrate that online sentiment influences the formation of offline public opinion; however, the relationships are not strong. The comparison between time-adjusted online sentiments and offline polls also confirms that the two public opinions are independent of each other regardless of time.

Public opinion from Twitter can represent specific groups in terms of age, gender, and political ideology. To test this possibility, this study uses subgroup information from

the polls and correlates the online opinions to these subgroups. Table 6 shows the correlation coefficients for the different age groups. Focusing on the relation between online sentiments and Realmeter, people in their 30s have the highest correlations at 0.185 and 0.233 for positive and negative sentiments, respectively. Meanwhile, those over 60 also have statistically significant coefficients compared with other age groups at 0.21 and 0.176, respectively. The 60-plus age group stands out in their correlations with the Gallup poll at 0.282 and 0.369, respectively.

**Table 6.** Correlation between datasets and the polls on different age groups.

| | | Realmeter | | Gallup | |
|---|---|---|---|---|---|
| | | **Positive** | **Negative** | **Positive** | **Negative** |
| Online Sentiments | 20 | 0.100 (0.326) | 0.240 (0.271) | 0.266 (0.348) | 0.071 (0.395) |
| | 30 | 0.185 (0.338) | 0.233 (0.321) | 0.167 (0.402) | 0.123 (0.379) |
| | 40 | 0.116 (0.501) | 0.212 (0.428) | 0.030 (0.406) | 0.065 (0.392) |
| | 50 | 0.014 (0.393) | 0.156 (0.368) | −0.012 (0.379) | 0.048 (0.331) |
| | 60+ | 0.210 (0.488) | 0.176 (0.258) | 0.282 (0.340) | 0.369 (0.256) |

Standard errors in parentheses.

The general notion is that the younger generation actively uses SNS, yet this result reveals a different story: older people may also passionately express their thoughts on political issues through SNS. Gender reveals an intriguing result as well. The online sentiments are closely related to Realmeter for males and Gallup for females. Table 7 shows the detailed results for different gender groups.

**Table 7.** Correlation between datasets and the polls on gender.

| | | Realmeter | | Gallup | |
|---|---|---|---|---|---|
| | | **Positive** | **Negative** | **Positive** | **Negative** |
| Online Sentiments | Male | 0.302 (0.562) | 0.411 (0.484) | 0.191 (0.606) | 0.195 (0.499) |
| | Female | 0.063 (0.479) | 0.156 (0.343) | 0.320 (0.509) | 0.283 (0.455) |

Standard errors in parentheses.

With regard to Realmeter, online sentiment has the highest correlation with the male group: 0.302 for positive and 0.411 for negative. These are significantly higher than the other values except for Gallup's online sentiments among females: 0.32 and 0.283, respectively. The correlation values for these gender groups are notably higher than for the other possible combinations.

Table 8 shows the correlation coefficients between online attitudes and political ideologies. There are three political ideology groups in South Korea: conservative, progressive, and neutral.

The neutral group from the Realmeter poll has outstanding values compared with the others: 0.310 for positive and 0.327 for negative. Other political ideology groups in the same poll do not reveal any relations to online sentiments. Gallup's conservative group shows some correlation, but this is lower than the simple one-to-one comparison.

**Table 8.** Correlation between datasets and polls on political ideology.

| | | Realmeter | | Gallup | |
|---|---|---|---|---|---|
| | | **Positive** | **Negative** | **Positive** | **Negative** |
| Online Sentiments | Conservative | −0.127 (0.559) | 0.015 (0.466) | 0.135 (0.408) | 0.198 (0.322) |
| | Neutral | 0.310 (0.317) | 0.327 (0.280) | 0.094 (0.432) | 0.160 (0.371) |
| | Progressive | −0.263 (0.247) | −0.060 (0.373) | 0.152 (0.479) | 0.059 (0.492) |

Standard errors in parentheses.

The analysis shows that online sentiments do not correlate to offline polls. Public opinion extracted from Twitter provides a different and independent trend compared to existing measures of presidential approval. In a small time-window, online and offline sentiments may be similar, but they are not closely related for longer time frames. Transforming online sentiments closer to the polls by reducing volatility increases the correlation in a limited manner. In addition, shifting online sentiments before and after the given date of the polls does not significantly increase the correlation. The time differential yields mixed results between Gallup and Realmeter. Online presidential approval is not a subset of offline ratings. All combinations of online and offline public opinion have only weak correlations in terms of gender, age, and political ideology. The analysis consistently indicates that presidential approval measured from Twitter is not substantially associated with offline polls. The results imply that public opinions online can represent the independent population as opposed to offline surveys.

## 5. Conclusions

This study investigates a method for measuring aggregated sentiment in cyberspace and explores the characteristics of online sentiments by comparing them to offline polls. It uses supervised deep learning to extract user attitudes from text and translates measured sentiments into the public opinion of people online. The study emphasizes that the deep learning model processes non-English user-generated data for sentiment analysis and its application to politics. Presidential approval is the most popular and the most studied public opinion in the field of political science. Many studies have been conducted to understand its effects and determinants [50]. The present study analyzes presidential approval by comparing and contrasting online ratings with offline ones. Evaluating online public opinion involves three stages: first, a machine collects footprints of people from the Internet. The massive amount of online textual data necessitates the use of a computer. Second, human coders label a text with the appropriate attitude. This process allows a machine to learn how to determine sentiments like a person. Third, deep learning algorithms study human-coded text–sentiment pairs and determine sentiments for all collected texts. The study calculates accuracy for all combinations using CNN, RNN, and BERT with different embedding types and sizes.

This study finds that the best-fitting algorithm is the modified BERT model, from which aggregated online sentiment is obtained. The trained algorithm yields a sentiment prediction accuracy of 94.18%, which is better than the rate at which the coders unanimously determine sentiment for the prepared datasets (78.45%). This method illustrates the possibility of collecting and translating unstructured data into a suitable form to use in political science research. With proper data processing, a computer algorithm can extract sentiment from plain text. Text–sentiment becomes a group attitude—in other words, public opinion—when they are aggregated accordingly.

In addition, the study finds that online sentiments are different from offline polls. Specifically, online sentiments toward a president do not correlate with conventional presidential approval ratings. Online public opinion is much more volatile and instantaneous

than offline. Weekly and monthly transformations, which are types of noise reduction, improve correlation in a limited manner. Although certain time adjustments slightly increase correlation, the results for polling agencies are mixed. Age and gender groups in offline polls are not significantly highly correlated with online sentiment. The minor increase in correlation depending on a comparison pair does not imply that online public opinion is a subset of offline polls.

This study demonstrates that one may measure groups' aggregated attitudes using unstructured data from the Internet with the help of deep learning. It also explains that public opinions from both online and offline environments are fundamentally different through various correlation analyses. Online sentiments exist parallel to public opinion as measured by polls. As people's engagement with the Internet continuously increases, they leave more clues about their thoughts and behaviors. Using these online traces has some implications that can broaden our understanding of people and society.

This study has several limitations and possible improvements for future research must be highlighted. First, the study only used user-generated Twitter texts. While this has advantages, for instance, easy access via API and abundance of subsidiary information except an actual text, it is one of many online platforms where online users reside. Therefore, a mixture of different online services may deliver online public opinion similar to that of offline. In addition, including other online channels will allow researchers to perform a comparative analysis of mass opinions from these different services. Second, the study includes Twitter texts from 2019. Considering the volatile political environment, future research can incorporate data from a longer period. Moreover, it is possible to divide an entire time into periods and analyze the differences between online and offline public opinion to measure the potential influence between the two mass sentiments. Third, this study explores the application of the method to analyze a political context, specifically presidential approval. It is possible to measure different issues such as gender and economy from cyberspace and illustrate characteristics of online public opinion through quantitative and qualitative research. Finally, public opinion studies attempt to discover important dependent and independent variables. Therefore, future research can investigate political factors influencing the mass online opinion and political outcomes affected by online public opinion including Twitter sentiments. This study can be considered as explorative research if deep learning techniques can complement political science by processing and generating relevant information from non-English unstructured data. Therefore, future research is necessary to improve the method for measuring online public opinion and understand its qualities to provide materials for more effective decision making.

The study focuses on how to measure online public opinion on a specific subject: the president of South Korea. This method can expand to various studies including those on environmental and cultural issues, which have greater online presence. It complements traditional polling by providing an abundance of data and greater anonymity, which help researchers better understand people's aggregated thoughts. Future research can test the feasibility of the method on various subjects and propose modifications depending on the peculiarity of an issue. Furthermore, scholars must analyze which public opinion—online or offline—is more important in decision-making processes to assess the practicality of the methods.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. MacDougall, C.D. *Understanding Public Opinion: A Guide for Newspapermen and Newspaper Readers*; The Macmillan Company: New York, NY, USA, 1952.
2. Gronke, P.; Newman, B. Public Evaluations of Presidents. In *The Oxford Handbook of the American Presidency*; Edwards, G.C., III, Howel, W., Eds.; Oxford University Press: New York, NY, USA, 2009; pp. 232–253.
3. Kreiss, D. Seizing the moment: The presidential campaigns' use of Twitter during the 2012 electoral cycle. *New Media Soc.* **2016**, *18*, 1473–1490. [CrossRef]
4. Klašnja, M.; Barberá, P.; Beauchamp, N.; Nagler, J.; Tucker, J. Measuring public opinion with social media data. In *The Oxford Handbook of Polling and Survey Methods*; Oxford University Press: Oxford, UK, 2017.
5. Tucker, J.A.; Nagler, J.; MacDuffee, M.M.; Barberá, P.; Penfold-Brown, D.; Bonneau, R. Big data, social media, and protest: Foundations for a research agenda. In *Computational Social Science: Discovery and Prediction*; Alvarez, R.M., Ed.; Cambridge University Press: New York, NY, USA, 2016; pp. 199–224.
6. Tufekci, Z.; Wilson, C. Social media and the decision to participate in political protest: Observations from Tahrir Square. *J. Commun.* **2012**, *62*, 363–379. [CrossRef]
7. Erikson, R.S.; Tedin, K.L. *American Public Opinion*; Routledge: New York, NY, USA, 2015.
8. Stimson, J.A. *Public Opinion in America: Moods, Cycles, and Swings*; Routledge: London, UK, 2018.
9. Gartner, S.S.; Segura, G.M. War, casualties, and public opinion. *J. Confl. Resolut.* **1998**, *42*, 278–300. [CrossRef]
10. Campbell, A.; Converse, P.E.; Miller, W.E.; Stokes, D.E. *The American Voter*; University of Chicago Press: Chicago, IL, USA, 1980.
11. Berinsky, A.J. Measuring public opinion with surveys. *Annu. Rev. Political Sci.* **2017**, *20*, 309–329. [CrossRef]
12. Koo, B. Automatic response system and its failure in sampling young female Korean voters: Gender gap in politically opinionated voters. *Surv. Res.* **2017**, *18*, 31–60. [CrossRef]
13. Canes-Wrone, B. *Who Leads Whom? The Policy Effects of Presidents' Relationship with the Masses*; University of Chicago Press: Chicago, IL, USA, 2005.
14. Eshbaugh-Soha, M.; Rottinghaus, B. Presidential position taking and the puzzle of representation. *Pres. Stud. Q.* **2013**, *43*, 1–15. [CrossRef]
15. Canes-Wrone, B.; De Marchi, S. Presidential approval and legislative success. *J. Politics* **2002**, *64*, 491–509. [CrossRef]
16. Mueller, J.E. *War, Presidents, and Public Opinion*; John Wiley & Sons: Hoboken, NJ, USA, 1973.
17. Berlemann, M.; Enkelmann, S. The economic determinants of US presidential approval: A survey. *Eur. J. Political Econ.* **2014**, *36*, 41–54. [CrossRef]
18. MacKuen, M.B.; Erikson, R.S.; Stimson, J.A. Peasants or bankers? The American electorate and the US economy. *Am. Political Sci. Rev.* **1992**, *86*, 597–611.
19. Norpoth, H. Presidents and the prospective voter. *J. Politics* **1996**, *58*, 776–792. [CrossRef]
20. Ostrom, C.; Simon, D. Promise and Performance: A Dynamic Model of Presidential Popularity. *Am. Political Sci. Rev.* **1985**, *79*, 334–358. [CrossRef]
21. Duggan, M.; Brenner, J. *The Demographics of Social Media Users, 2012*; Pew Research Center's Internet & American Life Project: Washington, DC, USA, 2013; Volume 14.
22. Mellon, J.; Prosser, C. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Res. Politics* **2017**, *4*, 2053168017720008. [CrossRef]
23. Eom, K. Whose Opinions Are Represented in the Online World? *J. Future Politics* **2020**, *10*, 99–122.
24. Lee, H.; Yang, S.M. Effects of demographics and personality factors on categorizing SNS users into three groups: Non-user, light user, and heavy users. *Korean J. Broadcast. Telecommun. Stud.* **2017**, *31*, 5–40.
25. Barberá, P.; Rivero, G. Understanding the political representativeness of Twitter users. *Soc. Sci. Comput. Rev.* **2015**, *33*, 712–729. [CrossRef]
26. Vaccari, C.; Valeriani, A.; Barberá, P.; Bonneau, R.; Jost, J.T.; Nagler, J.; Tucker, J. Social media and political communication. A survey of Twitter users during the 2013 Italian general election. *Riv. Ital. Di Sci. Politica* **2013**, *43*, 381–410.
27. Hampton, K.N.; Goulet, L.S.; Rainie, L.; Purcell, K. *Social Networking Sites and Our Lives*; Pew Internet & American Life Project: Washington, DC, USA, 2011; Volume 1.
28. Franch, F. 2010 UK Election prediction with social media. *J. Inf. Technol. Politics* **2013**, *10*, 57–71. [CrossRef]
29. Jensen, M.J.; Anstead, N. Psephological investigations: Tweets, votes, and unknown unknowns in the republican nomination process. *Policy Internet* **2013**, *5*, 161–182. [CrossRef]
30. Tumasjan, A.; Sprenger, T.O.; Sandner, P.G.; Welpe, I.M. Election forecasts with Twitter: How 140 characters reflect the political landscape. *Soc. Sci. Comput. Rev.* **2011**, *29*, 402–418. [CrossRef]
31. Elmer, G. Live research: Twittering an election debate. *New Media Soc.* **2013**, *15*, 18–30. [CrossRef]

32. Shamma, D.A.; Kennedy, L.; Churchill, E.F. Tweet the debates: Understanding community annotation of uncollected sources. In Proceedings of the First SIGMM Workshop on Social Media ACM, Beijing, China, 23 October 2009; pp. 3–10.

33. Gonzalez-Bailon, S.; Banchs, R.E.; Kaltenbrunner, A. Emotional reactions and the pulse of public opinion: Measuring the impact of political events on the sentiment of online discussions. *arXiv* **2010**, arXiv:1009.4019.

34. Chung, C.; Rhee, Y.; Cha, H. Big Data analyses of Korea's nation branding on Google and Facebook. *Korea Obs.* **2020**, *51*, 151–174. [CrossRef]

35. Skoric, M.; Poor, N.; Achananuparp, P.; Lim, E.P.; Jiang, J. Tweets and votes: A study of the 2011 Singapore general election. In Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2012; pp. 2583–2591.

36. Bermingham, A.; Smeaton, A. On using Twitter to monitor political sentiment and predict election results. In Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP 2011), Chiang Mai, Thailand, 13 November 2011; pp. 2–10.

37. Liu, B. *Sentiment Analysis and Opinion Mining*; Morgan & Claypool: Chicago, IL, USA, 2012.

38. Alaparthi, S.; Mishra, M. BERT: A sentiment analysis odyssey. *J. Mark. Anal.* **2021**, *9*, 118–126. [CrossRef]

39. Dodds, P.S.; Harris, K.D.; Kloumann, I.M.; Bliss, C.A.; Danforth, C.M. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PLoS ONE* **2011**, *6*, e26752. [CrossRef] [PubMed]

40. González-Bailón, S.; Banchs, R.E.; Kaltenbrunner, A. Emotions, public opinion, and US presidential approval rates: A 5-year analysis of online political discussions. *Hum. Commun. Res.* **2012**, *38*, 121–143. [CrossRef]

41. Thelwall, M.; Buckley, K.; Paltoglou, G.; Cai, D.; Kappas, A. Sentiment strength detection in short informal text. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *62*, 419. [CrossRef]

42. Young, L.; Soroka, S. Affective news: The automated coding of sentiment in political texts. *Political Commun.* **2012**, *29*, 205–231. [CrossRef]

43. Korea Information Society Development Institute. Analysis on SNS Trend and Use Behavior. KISDI STAT Report 19-10. 2019. Available online: https://stat.kisdi.re.kr/MediaStat/Library/Library_detail1.aspx?MENU_ID=233&Division=1&seq=2466 (accessed on 28 February 2019).

44. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.

45. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.

46. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

47. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]

48. Vaswani, A.; Bengio, S.; Brevdo, E.; Chollet, F.; Gomez, A.N.; Gouws, S.; Jones, L.; Kaiser, Ł.; Kalchbrenner, N.; Parmar, N.; et al. Tensor2tensor for neural machine translation. *arXiv* **2018**, arXiv:1803.07416.

49. SKTBrain. Korean BERT Pre-Trained Cased (KoBERT). Available online: https://github.com/SKTBrain/KoBERT/ (accessed on 28 February 2020).

50. Gronke, P.; Koch, J.; Wilson, J.M. Follow the leader? Presidential approval, presidential support, and representatives' electoral fortunes. *J. Politics* **2003**, *65*, 785–808.