Article

SocialTERM-Extractor: Identifying and Predicting Social-Problem-Specific Key Noun Terms from a Large Number of Online News Articles Using Text Mining and Machine Learning Techniques

Jong Hwan Suh

Department of Management Information Systems, BERI, Gyeongsang National University, 501 Jinjudae-ro Jinju-si, Gyeongsangnam-do 52828, Korea; jonghwan.suh@gnu.ac.kr; Tel.: +82-55-772-1537

Received: 11 November 2018; Accepted: 25 December 2018; Published: 2 January 2019

Abstract: In the digital age, the abundant unstructured data on the Internet, particularly online news articles, provide opportunities for identifying social problems and understanding social systems for sustainability. However, the previous works have not paid attention to the social-problem-specific perspectives of such big data, and it is currently unclear how information technologies can use the big data to identify and manage the ongoing social problems. In this context, this paper introduces and focuses on social-problem-specific key noun terms, namely SocialTERMs, which can be used not only to search the Internet for social-problem-related data, but also to monitor the ongoing and future events of social problems. Moreover, to alleviate time-consuming human efforts in identifying the SocialTERMs, this paper designs and examines the SocialTERM-Extractor, which is an automatic approach for identifying the key noun terms of social-problem-related topics, namely SPRTs, in a large number of online news articles and predicting the SocialTERMs among the identified key noun terms. This paper has its novelty as the first trial to identify and predict the SocialTERMs from a large number of online news articles, and it contributes to literature by proposing three types of text-mining-based features, namely temporal weight, sentiment, and complex network structural features, and by comparing the performances of such features with various machine learning techniques including deep learning. Particularly, when applied to a large number of online news articles that had been published in South Korea over a 12-month period and mostly written in Korean, the experimental results showed that Boosting Decision Tree gave the best performances with the full feature sets. They showed that the SocialTERMs can be predicted with high performances by the proposed SocialTERM-Extractor. Eventually, this paper can be beneficial for individuals or organizations who want to explore and use social-problem-related data in a systematical manner for understanding and managing social problems even though they are unfamiliar with ongoing social problems.

Keywords: social-problem-specific key noun terms; temporal weights; sentiment analysis; complex network structure analysis; deep learning; ensemble learning methods

1. Introduction

1.1. Social Problems and Challenging Issues for Identifying Ongoing Social Problems

Facing social problems as the challenges of the well-being and sustainability, e.g., high suicide rate and air pollution with fine dust in South Korea, research and development (R&D) projects have been promoted in addition to political and administrative measures not only to solve social problems, and but also to improve the quality of people’s lives by solving social problems. Subsequently,
technologies for solving social problems have attracted increased attention and driven the expansion of social entrepreneurship around the world. Thus, both the public and private sectors have started to place more emphasis than before on the technologies that are related to social problems [1–3]. Moreover, technological knowledge shares, e.g., patent documents and open research articles, have been more easily accessible currently than ever before, and this provides opportunities for solving social problems based on technologies rather than government policies. Therefore, it is necessary to facilitate the exploration of technological knowledge shares for solving social problems, and we should focus on identifying the ongoing social problems because it is a starting point for linking social problems to state-of-the-art technologies as solutions.

At the same time, with the emergence of Web 2.0 and social media, the amount of unstructured, textual data on the Internet has grown tremendously, especially at the micro level, which involves human behaviors, such as tweets of an individual person, the textual expression in blog posts over a period time, and furthermore the sensed activities by digital sensors in the so-called Internet of Things [4]. This abundance of publicly available textual data creates new opportunities for both qualitative and quantitative researchers in various areas that are related to data and information sciences. In particular, the big data that are available on the Internet provide opportunities for identifying social problems. A large amount of event-related textual data, e.g., online news articles and web forum posts, contain data and information that are useful for taking an overview of ongoing social problems, e.g., the types of ongoing social problems and their progress. In this respect, several information technologies can be applied, from information retrieval (IR) to text mining. However, the previous works have not paid attention to social-problem-specific perspectives of big data, so it is currently unclear how information technologies can be used to identify and manage the ongoing social problems from big data. In detail, the following challenging issues need to be resolved:

First, various social problems are in process simultaneously, and they occur in multiple streams of events. Therefore, it is a nontrivial task to determine the landscape of ongoing social problems from a large amount of the event-related textual data, e.g., online news articles and web forum posts. Particularly when individual persons are unfamiliar with the ongoing social problems, it is difficult for them to identify the ongoing social problems from such big data [5]. Therefore, it is necessary to identify the ongoing social problems and their key terms, which can represent the ongoing social problems, in an automatic way.

Second, most people use nouns as key terms to get data and information about ongoing social problems because they are relevant to topics, whereas the other types of key terms such as verbs, adjectives, and adverbs are more relevant to sentiments than topics [6–8]. In addition, for the same reason, it is harder for people to come up with and figure out key noun terms that are related to social problems than the other types of key terms. It means key noun terms are crucial for successfully getting data and information about the ongoing social problems. Hence, we need to focus on key noun terms for finding out social-problem-related data and information.

Third, some of the key noun terms can be more useful for figuring out the topics of ongoing social problems from big data because they play roles in categorizing those social-problem-related topics (SPRTs) into their corresponding social problems. For example, let us assume that two SPRTs were detected, and each SPRT was represented by five key noun terms, namely topic1 = {Suwon, Suicide, Student, Female, Police} and topic2 = {Daegu, Suicide, Student, Male, Violence}. Then, the key noun terms, such as Suicide, Student, Female/Male, and Violence, can be social-problem-specific key noun terms (hereinafter, SocialTERMs). Particularly, Suicide indicates that the two SPRTs can be grouped into the same type of social problems. In contrast, the city names, namely Suwon and Daegu, are event-specific key noun terms (hereinafter, EventTERMs), which specify that the two SPRTs occurred at different locations. Thus, the different roles of such key noun terms in labeling the identified SPRTs need to be considered.

Lastly, there have been no previous works on identifying the SocialTERMs from a large number of online news articles in an automatic way. Consequently, manual annotation after reading a large
amount of textual data is unavoidable for now in extracting the SocialTERMs for the detected SPRTs. However, it requires significant human efforts, and it is labor-intensive, expensive, time-consuming, and often error-prone. In addition, to reflect the importance of key noun terms in the textual data at different levels, e.g., the document level and the detected topic level, several weighting schemes have been used in previous works, e.g., \(tf\), \(idf\), and \(tfidf\) [9]. However, it is unknown whether those weighting schemes can reflect the different roles of key noun terms in representing social problems over time.

1.2. Key Term Identification in the Previous Text Mining Applications

With the emergence of Web 2.0 and social media, the amount of unstructured data, most of which are textual and publicly available on the Internet, has increased massively, especially the amount of data on individual entities such as persons and companies. This big data creates new opportunities for both qualitative and quantitative researchers of data and information sciences. Thus, big data is essential not only for scientific research on social systems but also for businesses and individuals [10], and it is important to develop a method that helps people obtain the relevant data quickly and accurately from big data and analyze it. To resolve it, text mining has been employed, with a focus on analyzing the statistical properties of terms [11]. In particular, key terms are essentials for exploring the overall data set, and text mining uses them to inspect and process the obtained texts through typical preparation steps.

Table 1 shows recent studies (2014–2018) in which key terms were extracted and used for text mining applications. The previous works in Table 1 can be grouped according to the final application of their identified key terms: indexing, clustering, summarization, classification (or categorization), or mapping [12,13]. Indexing, in which textual data are represented with a set of extracted key terms, is a research goal on its own, and is also a step in most text mining applications, such as feature generation and text representation [14]. Clustering is to group textual data on the basis of their attributes to identify important themes, patterns, or trends [15], and it is employed for topic detection (TD) [16]. Summarization focuses on creating a summary that contains the most important points of the original documents [17]. Classification assigns textual data to two or more categories [18,19]. Mapping focuses on information visualization and supports effective and efficient searches of important subjects or topic areas, which are identified from textual data [15,20].

In addition, Table 1 provides three taxonomies for the key term identification of the previous works. First, the key terms can be discovered to best describe the textual data at different levels: the sentence level [21–23], the document level [24], and the topic level [6]. Second, the key term identifications used in the previous text mining applications can be divided into three categories: manual, automatic, and hybrid approaches. Third, particularly for the automatic approach, four types of techniques have been used: statistical, linguistic, machine-learning-based, and hybrid approaches [12,24,25]. While the statistical approaches do not require any learning mechanism and use statistical information of terms, e.g., \(tf\), \(idf\), and \(tfidf\) [26,27], the linguistic approaches use linguistic features of terms, e.g., parsing, sentiment analysis, and semantics [28,29]. Machine-learning-based approaches use key terms that are extracted from the collected textual data by means of a training process and apply them to a machine learning model to find key terms in new textual data [12,28]. The hybrid approaches combine one or more techniques [17,30].
Table 1. Recent works (2014–2018) that extract and use key terms for text mining applications.

| Final Goals of Using Key Terms | Previous Works with Description | Type of Data | Level of Textual Data Analysis | Category of Key Term Extraction | Technique Used for Key Term Extraction |
|--------------------------------|--------------------------------|--------------|-------------------------------|---------------------------------|----------------------------------------|
|                                |                                |              | L1   | L2   | L3 | C1 | C2 | C3 | T1 | T2 | T3 | T4 |
| Indexing                       | Abilhoa and de Castro [25] proposed a keyword extraction technique of Twitter messages in which tweets are represented as graphs. | Tweets       | ✓    | ✓    | *  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
|                                | Nob, Jo and Lee [26] explored strategies for selecting and processing keywords for patent analysis purposes. | Patent documents | ✓    | ✓    | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
|                                | Almeida et al. [31] proposed a method to normalize and expand original short and messy text messages. | Short message service (SMS) | ✓    | ✓    | *  | ✓  | ✓  | *  | ✓  | ✓  | ✓  | ✓  |
|                                | Rao et al. [32] presented a new term weighting scheme called LGT, which jointly models the Local element, Global element, and Topical association of each story. | Online news articles | ✓    | ✓    | ✓  | *  | ✓  | *  | ✓  | ✓  | ✓  | ✓  |
|                                | Lin et al. [33] proposed an Explicit Emotion Signal based cross media sentiment learning approach. | Microblog (i.e., Sina Weibo) posts | ✓    | ✓    | ✓  | *  | ✓  | *  | ✓  | ✓  | ✓  | ✓  |
| Clustering                     | Jiang, Chen, Nunamaker and Zimbra [6] proposed a novel stakeholder-based event analysis framework that uses online stylometric analysis, and partitions their messages into different time periods of major firm events. | Web forum posts | ✓    | ✓    | *  | ✓  | ✓  | *  | ✓  | ✓  | ✓  | ✓  |
|                                | Alruily et al. [34] examined the crime domain in the Arabic language (unstructured text) using text mining techniques, and presented the development and application. | Online news articles | ✓    | ✓    | *  | ✓  | ✓  | *  | ✓  | ✓  | ✓  | ✓  |
|                                | Lo et al. [35] presented an unsupervised multilingual approach for identifying highly relevant terms and topics from the mass of social media data. | Tweets        | ✓    | ✓    | ✓  | *  | ✓  | *  | ✓  | ✓  | ✓  | ✓  |
|                                | Pournarakis et al. [36] devised a novel genetic algorithm to improve clustering of tweets in semantically coherent groups. | Tweets        | ✓    | ✓    | ✓  | *  | ✓  | *  | ✓  | ✓  | ✓  | ✓  |
| Summarization                  | Zhang et al. [37] presented six term clumping steps that can clean and consolidate topical content in text sources for tech mining. | Research articles | ✓    | ✓    | ✓  | *  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
|                                | Zheng, Lin, Wang, Lin and Song [23] studied an approach to extract product and service aspect words, as well as sentiment words, automatically from reviews. | Reviews       | ✓    | ✓    | ✓  | *  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
|                                | Li et al. [38] proposed a convolutional neural network (CNN)-based opinion summarization method for Chinese microblogging systems. | Microblog (i.e., Sina Weibo) posts | ✓    | ✓    | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
|                                | Hu, Chen and Chou [17] proposed a novel multi-text summarization technique for identifying the top-k most informative sentences of hotel reviews. | Reviews       | ✓    | ✓    | ✓  | *  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
Table 1. Cont.

| Final Goals of Using Key Terms | Previous Works with Description | Type of Data          | Level of Textual Data Analysis ¹ | Category of Key Term Extraction ² | Technique Used for Key Term Extraction ³ |
|-------------------------------|---------------------------------|-----------------------|----------------------------------|-----------------------------------|------------------------------------------|
|                               |                                 |                       | L1  | L2  | L3  | C1  | C2  | C3  | T1  | T2  | T3  | T4  |
| Classification                |                                 |                       |     |     |     |     |     |     |     |     |     |     |
| Weichselbraun et al. [39]     | presented a novel method for contextualizing and enriching large semantic knowledge bases for opinion mining with a focus on Web intelligence platforms and other high-throughput big data applications. | Reviews               | √   |     |     |     |     |     |     |     |     |     |
| Xu, Zhang and Wang [22]       | proposed a support vector machine (SVM)-based approach to identify implicit features from Chinese customer reviews. | Reviews               | √   |     |     | *   |     |     | *   |     | *   |     |
| Peetz, de Rijke and Kaptein [30] | proposed a feature-based model based on three dimensions, i.e., the source of the tweet, the contents of the tweet and the reception of the tweet. | Tweets                | √   |     |     | *   |     |     | *   |     | *   |     |
| Li and Liu [40]               | established a classification model that predicts the temporal class of an original microblog’s retweeting time series by using readily available social-influential, topical, and temporal factors. | Microblog (i.e., Sina Weibo) posts | √   |     |     |     |     |     |     |     |     |     |
| This study                    |                                 | Online news articles   | √   |     |     |     | *   | *   | *   |     |     |     |
| Mapping                       |                                 |                       |     |     |     |     |     |     |     |     |     |     |
| Jung and Segev [41]           | proposed methods to analyze how communities change over time in the citation network graph without additional external information and based on node and link prediction and community detection. | Research articles     | √   |     |     |     |     |     |     |     |     |     |
| Lee et al. [42]               | suggested a way of technology opportunity identification that is customizable to the R&D capabilities of small and medium-sized enterprises (SMEs). | Patent documents      | √   |     |     | *   |     |     | *   |     |     |     |
| Yang, Han, Wolfram and Zhao [28] | introduced the author keyword coupling analysis (AKCA) method to visualize the field of information science (2006–2015). | Research articles     | √   |     |     | *   |     | *   | *   |     | *   |     |
| Piryani, Madhavi and Singh [27] | presented a scientometric mapping of research work done on opinion mining and sentiment analysis (OMSA) during 2000–2016. | Research articles     | √   |     |     | *   |     | *   | *   |     | *   |     |

Notes: ¹ Level (L): sentence (L1), document (L2), and topic (L3). ² Category of key term extraction (C): manual (C1), automatic (C2), and hybrid (C3). The asterisk * shows the previous works with * belong to the categories, checked with * and composing the hybrid type C3. ³ Techniques used for key term extraction (T): statistical (T1), linguistics (T2), machine learning (T3), and hybrid (T4). The asterisk * shows the previous works with * used the technique(s), checked with * and composing the hybrid type T4.
According to Table 1, most prior text mining applications are either automatic approaches or hybrid approaches, and they have adopted either statistical techniques or hybrid techniques for extracting their key terms. Theoretically, under these taxonomies, this study can be categorized into classifying the topic-level key noun terms from online news articles into the SocialTERMs and the EventTERMs by adopting a hybrid technique, as highlighted in Table 1. Consequently, according to Table 1, the key theoretical contributions of this paper can be summarized as follows:

First, to the best of our knowledge, based on Table 1, there exists a research gap that no previous work has dealt with: the automatic classification of key noun terms that were identified by TD into the SocialTERMs and the EventTERMs. This paper contributes to addressing this research gap.

Second, to label the identified topics, most of the previous works in Table 1 used simple statistical approaches for characterizing key terms from clustered documents. On the other hand, this paper proposes and employs temporal weight features, sentiment features, and complex network structural features to represent key noun terms, which can be identified to label the detected SPRTs, after reviewing the features that were used in the previous works of Table 1.

Third, according to Table 1, no previous study has compared the performances of the state-of-the-art classification techniques, particularly deep learning, for distinguishing between the SocialTERMs and the EventTERMs among the key terms of the detected SPRTs. This paper extends the related literature by taking on such a challenging issue.

1.3. Purpose and Organization of This Paper

To resolve the abovementioned challenging issues, this paper proposes an automatic approach, namely SocialTERM-Extractor, which identifies the SPRTs from a large number of Korean online news articles and classifies the key noun terms of the detected SPRTs into the SocialTERMs and the EventTERMs.

To design and examine the proposed approach, three research questions can be formulated as below, and a research framework is constructed to answer those research questions:

RQ1. How well do the three types of features, namely temporal weight, sentiment, and complex network structural features, perform in distinguishing the SocialTERMs and the EventTERMs among the key noun terms of the detected SPRTs from a large number of online news articles by using different classification techniques? Moreover, which feature set and features give the best results?

RQ2. Which classification technique among the five base learners, namely Decision Tree (DT), Naïve Bayes (NB), Radial Basis Function Network (RBFN), Support Vector Machine (SVM), and Deep Belief Network (DBN), is best suited for differentiating the key noun terms of the detected SPRTs into the SocialTERMs and the EventTERMs?

RQ3. Which ensemble learning method gives the best results? Is there a single ensemble method that achieves the best performances for the all feature sets with any given base learner?

The rest of the paper is organized as follows: Section 2 outlines the proposed research framework to design and examine the SocialTERM-Extractor, and explains it in detail. Section 3 presents the results of applying the suggested research framework to the online news articles, which are collected from the best-known Korean news portal site in South Korea. Section 4 discusses the application results in terms of designing an automatic system. Finally, Section 5 presents the conclusions of this paper with reflections on limitations and future works.

2. Materials and Methods

To answer the research questions in the previous section, a research framework is proposed as summarized in Figure 1. First, online news articles reporting social-problem-related events are collected from a test-bed news portal site. Second, sentiment analysis selects online news articles with negative sentiment from the collected data. Then, from the online news articles having negative sentiment, the SPRTs are detected, and labelled by their key noun terms. Third, the three types
of features regarding the key noun terms of the detected SPRTs are measured: temporal weight, sentiment, and complex network structural features. Fourth, the configurations of different feature sets and different classification techniques are evaluated respect to three performance measures, namely accuracy, F-measure, and area under accuracy (AUC). Then, the comparative studies are performed for different feature sets and different classification techniques. Following subsections explain the steps of the proposed research framework in detail.

Figure 1. Research framework, proposed by this study to design and examine the SocialTERM-Extractor.

2.1. Collect Data

In the first component of Figure 1, news sections related to society are targeted for data collection. Then, the data collection is performed mainly by two steps, crawling and parsing:

First, a distributed web-crawling program is developed to collect the online news articles from the Internet in a significantly reduced timespan. In detail, the distributed web-crawling program is based on the simple remote procedure call (SRPC) framework, in which two tasks from a master computer are delivered to slave computers with various hardware configurations, i.e., a uniform resource identifier (URI) to crawl and how to crawl the given URI. Consequently, a large number of
online news articles, published in the chosen society-related news sections of a test-bed news portal service, are collected as raw HTML pages.

Second, the textual data in <title> ... </title> and <content> ... </content> of the collected online news articles are parsed out from the raw HTML pages, and are stored in a relational database. In addition, the publication date of each online news article is stored in the database for the TD. This results in NEWS$_{0,t} = \{\text{news} \mid \text{online news articles, published at time } t \text{ and collected from the chosen society-related news sections of the test-bed news portal service} \}$ for time $t = 1, \ldots, T$.

2.2. Detect Social-Problem-Related Topics (SPRTs)

2.2.1. Select Online News Articles with Negative Sentiment

In this step, online news articles with negative sentiment are selected to focus on online news articles more related to social problems. Considering the wide applicability to different languages, the sentiment of an online news article is obtained based on multilingual sentiment feature set, made by two main parts proposed in Dang et al. [43]: the extraction of English sentiment feature set from SentiWordNet (http://sentiwordnet.isti.cnr.it/), and the construction of multilingual sentiment feature set.

To explain, for the multilingual sentiment features in SentiWordNet, the average polarity score is calculated by using the prior–polarity formula, defined as

$$
score(senti, pos, pol) = \frac{\sum_{\text{synset} \in \text{SYNSET}(senti, pos, pol)} \text{swnscore}(\text{synset}, pos, pol)}{n(\bigcup_{\text{pol}} \text{SYNSET}(senti, pos, pol))},
$$

where $senti$ is a sentiment feature in SentiWordNet, $pos$ is a sentiment related part-of-speech (POS) sense, $pos \in \{\text{verb, adverb, adjective}\}$, $pol$ is a type of polarity scores, $pol \in \{\text{objective, positive, negative}\}$, $\text{SYNSET}(senti, pos, pol)$ is a set of synsets, i.e., synonyms, belonging to $senti$ when $pos$ and $pol$ are given, and $\text{swnscore}(\text{synset}, pos, pol)$ is the SentiWordNet score of $\text{synset}$ with given $pos$ and $pol$.

From $score(senti, pos, pol)$, the final sentiment score is determined by the sentiment feature–calculation strategy. In the strategy, the sentiment features satisfying both $score(senti, pos, pol = \text{objective}) < 0.5$ and $|score(senti, pos, pol = \text{negative})| \neq |score(senti, pos, pol = \text{positive})|$ are taken into account, and the final negative sentiment score of a multilingual sentiment feature is calculated as

$$
\text{finalnegscore}(senti, pos, pol) = \begin{cases} 
0 & \text{if } |score(senti, pos, pol = \text{negative})| < |score(senti, pos, pol = \text{positive})|, \\
|score(senti, pos, pol = \text{negative})| & \text{otherwise.}
\end{cases}
$$

Then, using the constructed multilingual sentiment feature set, the sentiment score of an online news article, $news$, is obtained by

$$
\text{newsnegscore}(news) = \frac{\sum_{pos} \sum_{senti \in \text{NEWSSENTI}(news)} \text{finalnegscore}(senti, pos)}{n(\text{NEWSSENTI}(news))},
$$

where $\text{NEWSSENTI}(news)$ is a set of the multilingual sentiment features appearing in $news \in NEWS_{0,t}$.

In particular, this study uses the multilingual sentiment feature set of English and Korean, constructed by Suh [8], because of the following reasons: it is based on the commonly used approach for measuring multilingual sentiment, proposed by Dang, Zhang and Chen [43]; it enables researchers of the other lingual cultures to make use of this study’s research framework; Korean, selected for this study, is taken into account as the non–English for the multilingual sentiment features; the sentiments of synonyms for a English sentiment feature are considered for the corresponding Korean sentiment feature; and the additional Korean sentiment features are generated and included to consider the negation.

To explain briefly, the Korean sentiment feature, constructed by Suh [8], inherits the final sentiment score and POS sense of its corresponding English sentiment feature, generated by Dang, Zhang
and Chen [43]. For instance, as shown in Table A1 of Appendix A, the sentiment value of ‘디럼 히/pv�+더/ef’ is −0.7500 for pos = verb, and comes from the corresponding English sentiment feature, ‘soil’. If the English sentiment feature has synonyms, the final sentiment scores of the synonyms are averaged for the corresponding Korean sentiment feature. For example, the sentiment value of ‘즉시/ma그’ is the average of sentiment values from five English sentiment features: ‘instantly’, ‘straight_away’, ‘right_away’, ‘at_once’, and ‘swiftly’. Moreover, if the morphological analysis splits the Korean sentiment feature into a stem and ending(s), the extended Korean sentiment features are generated by adding various endings to the stem in possible POS senses and tenses. For instance, the stem of ‘디럼 히/pv�+더/ef’ is ‘디럼 히/pv�’, and the extended Korean sentiment features of ‘디럼 히/pv�+더/ef’ are listed in Table A2 of Appendix A. They inherit sentiment values from the original Korean sentiment feature, ‘디럼 히/pv�+더/ef’, and, when negation is added to their endings, −1 is multiplied to their sentiment values.

As a consequence, to select online news articles more concerned with social problems, the online news articles with newsnegscore(news) > 0 are chosen for the TD of the next step. This leads to NEWS_{1,t} = \{\text{news} \mid \text{online news articles, published at time } t, \text{ and turned out to have negative sentiment’s online news articles from } NEWS_{0,t}\} \text{ for time } t = 1, \ldots , T.  

2.2.2. Detect the SPRTs from the Collected and Negative Online News Articles

An event is defined as a real-world incident that is related to time(s) and location(s), e.g., 9/11 attacks of 2001, Hurricane Catalina of 2004, and North Korea’s nuclear weapon test [44]. Due to the rapid growth and popularity of the Web, when an event occurs, a large number of event-related textual data are published online [45]. Generally, online news articles are starting points, and Web 2.0 has recently led to the tremendous distribution of online news articles through individuals on social media [46,47]. As a result, managing, interpreting, and analyzing such a huge volume of online news articles that are related to events has been a difficult task. To address this, many online news articles that are related to a set of events and interconnected with one another need to be grouped into the same topic [16]. Then, such topics and their changes can be identified over time by using TD methods [48]. Formally, a topic is a seminal event that is associated with all related events, that is, a set of related events [49].

Therefore, to detect the topics of this study’s interest, i.e., SPRTs, this step clusters online news articles, which are collected and evaluated to have negative sentiment. First, noun terms are identified from the online news articles through a series of natural language processing (NLP) techniques, i.e., spacing, part-of-speech (POS) tagging, regular expressions-based noun extraction, and stop words removal. In this way, only noun terms are used for the TD because of the following reasons: first, the target key terms of this study, i.e., SocialTERMs and EventTERMs, are noun terms according to the introduction of this paper; second, the other types of key terms, i.e., verbs, adjectives, and adverbs, are more relevant to sentiments rather than topics [6–8]. Next, let news be an online news article in NEWS_{1,t}, and noun be a noun term in NEWSNOUN_{0(news)} = \{\text{noun} \mid \text{all noun terms in news}\}. Then, the weight score of noun in news \in NEWS_{1,t} is obtained by

$$w(noun, news) = tf(noun, news) \times idf_t(noun) \times ths(noun, news).$$ (4)

Here, \(tf(noun, news)\) is the normalized frequency of noun appearing in news, and it is defined as

$$tf(noun, news) = \frac{f(noun, news)}{\max_{noun \in \text{NOUN}(news)}(f(noun, news))}.$$ (5)
where \( f(noun) \) is the frequency of \( noun \) in <content> ... </content> of \( news \). In addition, \( idf_t(noun) \) is the inverse document frequency of \( noun \), defined as

\[
idf_t(noun) = \log \left( \frac{H_t}{h_t(noun)} \right),
\]

(6)

where \( h_t(noun) \) is the number of online news articles containing \( noun \) among online news articles in \( NEWS_{1,t} \), and \( H_t \) is the number of online news articles in \( NEWS_{1,t} \). On the other hand, \( ths(noun, news) \) is the existence of \( noun \) in <title> ... </title> of \( news \), given by

\[
ths(noun, news) = \begin{cases} 
1 & \text{if } noun \text{ appears in } <title> ... </title> \text{ of } news, \\
0.5 & \text{otherwise.} 
\end{cases}
\]

(7)

Using the obtained \( w(noun, news) \) values, five noun terms with the highest weights are selected as key noun terms for \( news \). This results in \( NEWSNOUN_{1}(news) = \{ noun \mid \text{five key noun terms for } news \} \), and \( news \) is represented by the vector of its five key noun terms due to its simplicity, compared to the other textual representations models, e.g., the graph-based model, and the fuzzy set model [50,51]. Here, one may argue about how to decide the number of key noun terms for an online news article, but this study refers to the number of key noun terms used to represent an online news article in the previous works, i.e., three to five keywords [6,8,52,53]. Therefore, in this study, the number of key noun terms for an online news article is set to five as default.

Then, Algorithm 1 is adopted to cluster online news articles in \( NEWS_{1,t} \) for \( t = 1, \ldots, T \). The Algorithm 1 is a modified version from the algorithm that was used in He, Chang, Lim and Banerjee [16] and Suh [8], which has been widely used and known effective for TD because it overcomes the following drawbacks. While previous TD models are broadly classified into two types, i.e., non–probabilistic and probabilistic [16], non-probabilistic models do not provide the number of topic clusters, and existing probabilistic models, especially latent Dirichlet allocation (LDA), seem to be overly complex for the TD problems.

Consequently, the Algorithm 1 extracts the topics of similar online news articles from \( NEWS_{1,t} \) for \( t = 1, \ldots, T \). Over the iterations in executing Algorithm 1, the centroid of each topic keeps less than \( \alpha \) key noun terms while excluding less important key noun terms. In consequence, Algorithm 1 results in \( TOPIC(topic) = \{ topic \mid \text{SPRTs detected from } NEWS_{1,t} \text{ for } t = 1, \ldots, T \} \), \( TOPICNEWS(topic) = \{ news \mid \text{online news articles, classified to } topic \in TOPIC \} \), and \( TOPICNOUN(topic) = \{ noun \mid \text{key noun terms in the centroid of } topic \in TOPIC \} \).
Algorithm 1 Detecting the SPRTs from online news articles in NEWS$_{1,t}$ ($t = 1, \ldots, T$).

Input: Online news articles in NEWS$_{1,t}$ and their noun score vectors, and threshold $\epsilon$
Output: TOPIC, TOPICNEWS(topic), and TOPICNOUN(topic)

1: for time $t = 1$ (i.e., the first publication date among online news articles of NEWS$_{1,t}$) to $t = T$ (i.e., the last publication date among online news articles of NEWS$_{1,t}$) do
2: select online news articles in NEWS$_{1,t}$;
3: if $t = 1$ and $n$(TOPIC) = 0 then
4: create a topic, set the online news article as a centroid of the new topic, and announce it;
5: else
6: for each online news article of NEWS$_{1,t}$ do
7: compute the cosine similarity of an online news article with the centroid of each topic in TOPIC, defined as
8: \[
     \text{sim}(\mathbf{v}_{\text{centroid}}, \mathbf{v}_{\text{news}}) = \frac{\mathbf{v}_{\text{centroid}} \cdot \mathbf{v}_{\text{news}}}{\|\mathbf{v}_{\text{centroid}}\| \|\mathbf{v}_{\text{news}}\|} \tag{8}
\]
9: if cosine similarity $> \text{threshold } \epsilon$ then
10: assign the new article to the nearest topic and update the centroid of the nearest topic with the new article by averaging their weight vectors;
11: else
12: create a new topic, assign the online news article to the new topic, and announce it;
13: end if
14: if the number of noun terms in the updated centroid $> \alpha$ then
15: keep only the top $\alpha$ noun terms with the highest weight scores as key noun terms;
16: end if
17: end for
18: end if
19: end for
20: select topics with $> \beta$ online news articles, and define them as the SPRTs.

2.3. Measure the Three Types of Features to Represent the Key Noun Terms of the SPRTs

To label the identified topics, most of the previous works in Table 1 used simple statistical approaches for characterizing key terms from clustered documents. In contrast, this paper proposes and employs temporal weight features, sentiment features, and complex network structural features to represent key noun terms, which can be identified to label the detected SPRTs, after reviewing the features that were used in the previous works of Table 1. Details of the proposed three features can be explained as follows:

**Temporal weight features.** Temporal IR attempts to consider not only relevance but also temporal correspondence based on the underlying temporal factor behind search intension. A relatively large number of key noun terms, i.e., queries, for information access have temporal information needs [54]. Hence, to represent the temporally changing importance of a key noun term in the identified topics, this study modifies the traditional weighting statistics, e.g., $tf$, $idf$, and $tfidf$, by taking into account time, which yields temporal weight features. In addition, basic statistics such as the mean, variance, and |skewedness| are measured for the temporal weight features, to consider the distributional characteristics over the given time period. Here, the absolute value of skewedness is to measure the shape of skewedness irrespective of whether it is skewed to the left/negative or to the right/positive.

**Sentiment features.** The sentiment features of a key noun term are measured by sentiment analysis on the large-scale online news articles. In general, sentiment analysis determines whether a textual data instance is objective or subjective and whether a subjective textual data instance contains positive or negative statements, and measures the sentiment value of a subjective textual data instance [55,56]. In this paper, the approach of Suh [8] that uses SentiWordNet as a lexicon is adopted to extract multilingual sentiment features and score their sentiment values mainly for two reasons: it
enables researchers in the other countries to use the research framework of this paper by constructing sentiment features with their own languages; and it takes into account the negations. In addition, this study exploits the basic statistics of a key noun term’s sentiment features to represent the distributional characteristics over the news and topics that contain the key noun term.

**Complex network structural features.** Using the co-occurrence relationships of the key noun terms as links, which are called co-news and co-topic links, the complex networks of the key noun terms are constructed, and their complex network structural properties are measured by referring to the standard measures of node centrality, i.e., the degree, closeness, and betweenness centralities [57–59], and used as features for this study. In addition, after specifying a boundary, such as identified SPRTs and detected topical communities, to the complex networks of the key noun terms, the basic statistics are measured to represent the distributions of a key noun term’s in-boundary network properties over the different SPRTs and topical communities.

Thus, in this Section 2.3, the proposed three types of features are measured for all the extracted key noun terms of the SPRTs, i.e., \( \text{noun} \in \cup_{\text{topic} \in \text{TOPIC}} \cup_{\text{TOPICnoun} \in \text{TOPICnoun(topic)}} \). The measured three types of features for the topic-level key noun term are used to decide automatically whether the topic-level key noun term is a SocialTERM or EventTERM in the next Section 2.4.

### 2.3.1. Measure the Temporal Weight Features of the SPRTs’ Key Noun Terms

The temporal weight features of \( \text{noun} \in \cup_{\text{topic} \in \text{TOPIC}} \cup_{\text{TOPICnoun} \in \text{TOPICnoun(topic)}} \), namely F1, are measured with four respects: \( \text{df}, \text{tf}, \text{ths}, \) and \( \text{idf} \). Moreover, these temporal weight features are measured with two different respects: at the news level and at the topic level.

First, the temporal weight features of \( \text{noun} \) at the news level are measured as follows: Given that \( \text{NOUNNEWS}(\text{noun}) \) is a set of online news articles, containing \( \text{noun} \) and published at time \( t \), \( \text{dfs}^{\text{score}}(\text{noun}, t) \) is the normalized number of online news articles in \( \text{NOUNNEWS}(\text{noun}) \), and it is given by

\[
dfscore_{1,t}(\text{noun}) = \frac{n(\text{NOUNNEWS}(\text{noun}))}{n(\cup_{\text{noun}} \text{NOUNNEWS}(\text{noun}))}. \tag{9}
\]

Given that \( \text{tf}(\text{noun}, \text{news}) \) is the frequency of \( \text{noun} \), which occurred in the content of \( \text{news} \in \text{NOUNNEWS}(\text{noun}) \), \( \text{tf}^{\text{score}}_{1,t}(\text{noun}) \) is obtained by normalizing \( \text{tf}(\text{noun}, \text{news}) \) by the number of online news articles in \( \text{NOUNNEWS}(\text{noun}) \), and it is defined as

\[
tf^{\text{score}}_{1,t}(\text{noun}) = \frac{\sum_{\text{news} \in \text{NOUNNEWS}(\text{noun})} \text{tf}(\text{noun}, \text{news})}{n(\cup_{\text{noun}} \text{NOUNNEWS}(\text{noun}))}. \tag{10}
\]

\( \text{ths}^{\text{score}}_{1,t}(\text{noun}) \) is the normalized \( \text{ths}(\text{noun}, \text{news}) \) for online news articles in \( \text{NOUNNEWS}(\text{noun}) \), and it is given by

\[
titlescore_{1,t}(\text{noun}) = \frac{\sum_{\text{news} \in \text{NOUNNEWS}(\text{noun})} \text{ths}(\text{noun}, \text{news})}{n(\cup_{\text{noun}} \text{NOUNNEWS}(\text{noun}))}. \tag{11}
\]

where \( \text{ths}(\text{noun}, \text{news}) \) is 2 if \( \text{noun} \) appears in the title of \( \text{news} \), otherwise it is 1.

\( \text{idf}^{\text{score}}_{1,t}(\text{noun}) \) is the inverse of the number of online news articles, containing \( \text{noun} \) at time \( t \), and it is defined as

\[
\text{idf}^{\text{score}}_{1,t}(\text{noun}) = \log\left(\frac{n(\cup_{\text{topic}} \text{TOPICNEWS}(\text{topic}))}{n(\cup_{\text{noun}} \text{NOUNNEWS}(\text{noun}))}\right). \tag{12}
\]

To represent the distribution of each of the Equations (9)–(12) over time \( t = 1, \ldots , T \), the mean, variance, and \( \text{skewness} \) are measured, and added as the temporal weight features of \( \text{noun} \) at the news level to F1. As a consequence, 12 features are measured as the news-level temporal weight features of \( \text{noun} \).
Second, the temporal weight features of noun at the topic level are obtained as follows: Given that NOUNTOPIC(noun) is a set of the detected SPRTs that contain online news articles in NOUNNEWS(noun) and thereby are related to noun, dfscore2,t(noun) is the normalized number of the detected SPRTs in NOUNTOPIC(noun), defined as

\[
dfscore2,t(noun) = \frac{n(NOUNTOPIC(noun))}{n(\cup_n NOUNTOPIC(noun))}.
\]  

(13)

tfscore2,t(noun) is obtained by normalizing tfscore1,t(noun) over the detected SPRTs related to noun, and defined as

\[
tfscore2,t(noun) = \frac{tfscore1,t(noun)}{n(NOUNTOPIC(noun))}.
\]  

(14)

In the same way, titlescore2,t(noun) is got by normalizing titlescore1,t(noun) over the detected SPRTs related to noun, given by

\[
titlescore2,t(noun) = \frac{titlescore1,t(noun)}{n(NOUNTOPIC(noun))}.
\]  

(15)

and idfscore2,t(noun) is the normalized idfscore1,t(noun) over the detected SPRTs regarding to noun, given by

\[
idfscore2,t(noun) = \log\left(\frac{n(TOPIC_t(noun))}{n(NOUNTOPIC_t(noun))}\right).
\]  

(16)

To represent the distribution of each of Equations (13)–(16) over time \(t = 1, \ldots, T\), 12 topic-level temporal weight features on noun are measured, and added to F1. Consequently, Table 2 shows that 24 temporal weight features on noun, measured at both news and topic levels.

Table 2. Temporal weight features of the detected SPRTs’ key noun terms, proposed for this study.

| Feature Sub Set | Level | Temporal Weight Feature ¹ |
|-----------------|-------|---------------------------|
| F11             | news  | mean, variance, and | skewness | of dfscore1,t(noun) |
|                 | topic | mean, variance, and | skewness | of dfscore2,t(noun) |
| F12             | news  | mean, variance, and | skewness | of tfscore1,t(noun) |
|                 | topic | mean, variance, and | skewness | of tfscore2,t(noun) |
| F13             | news  | mean, variance, and | skewness | of titlescore1,t(noun) |
|                 | topic | mean, variance, and | skewness | of titlescore2,t(noun) |
| F14             | news  | mean, variance, and | skewness | of idfscore1,t(noun) |
|                 | topic | mean, variance, and | skewness | of idfscore2,t(noun) |

Notes: ¹ The statistics of temporal weights of noun are obtained over \(t = 1, \ldots, T\).

2.3.2. Measure the Sentiment Features of the SPRTs’ Key Noun Terms

This component extracts features related to the sentiment of noun, namely F2. To do so, this paper adopts the multilingual sentiment feature set, constructed by Suh [8] through two main parts: the extraction of English sentiment feature set from SentiWordNet (http://sentiwordnet.isti.cnr.it/), and the construction of multilingual sentiment feature set.

Let NOUNSENTI(noun) be a set of the constructed multilingual sentiment features that contain noun. Then, the sentiment score of a multilingual sentiment feature including noun with the given pos is obtained by

\[
featuresentiscore(noun, pos) = \frac{\sum_{senti\in NOUNSENTI(noun)} finalscore(senti, pos)}{n(NOUNSENTI(noun))}.
\]  

(17)
In addition, the sentiment score of noun is defined as

\[
noun\text{sentiscore}(\text{noun}) = \frac{1}{3} \sum_{\text{pos}} \text{featuresenticore}(\text{noun, pos}).
\]  

(18)

The sentiment score of noun at the news level is obtained by averaging the sentiment scores of online news articles, containing noun, and it is given by

\[
sentiscore_{1}(\text{noun}) = \frac{\sum_{\text{news} \in \text{NOUNNEWS}(\text{noun})} \text{newssentiscore}(\text{news})}{n(\text{NOUNNEWS}(\text{noun}))},
\]  

(19)

where \text{NOUNNEWS}(\text{noun}) = \text{NOUNNEWS}_{1}(\text{noun}) \cup \ldots \cup \text{NOUNNEWS}_{T}(\text{noun}) for time \( t = 1, \ldots , T \).

Here, the sentiment score of an online news article, news, is given by

\[
\text{newssentiscore}(\text{news}) = \frac{\frac{1}{3} \sum_{\text{pos}} \sum_{\text{senti} \in \text{NEWSSEN}TI(\text{news})} \text{finalscore}(\text{senti, pos})}{n(\text{NEWSSEN}TI(\text{news}))},
\]  

(20)

where \text{NEWSSEN}TI(\text{news}) is a set of the multilingual sentiment features appearing in news. In addition, to represent the distribution of the sentiment scores over the detected SPRTs, whose online news articles are containing noun, the mean, variance, and \( |\text{skewness}| \) for \text{newssentiscore}(\text{news}) are measured over \text{news} \( \in \text{NOUNNEWS}(\text{noun}) \), and they are added as the sentiment features of noun to F2. Here, the mean value of \text{newssentiscore}(\text{news}) is equal to \text{sentiscore}_{1}(\text{noun})

The sentiment score of noun at the topic level is defined as

\[
sentiscore_{2}(\text{noun}) = \frac{\sum_{\text{topic} \in \text{NOTOUNTOPIC}(\text{noun})} \text{topicsentiscore}(\text{topic})}{n(\text{NOTOUNTOPIC}(\text{noun}))}.
\]  

(21)

Here, the sentiment score of the detected topic, topic, is given by

\[
\text{topicsentiscore}(\text{topic}) = \frac{\sum_{\text{news} \in \text{TOPICNEWS}(\text{topic})} \text{newssentiscore}(\text{news})}{n(\text{TOPICNEWS}(\text{topic}))}.
\]  

(22)

In addition, to represent the distribution of the sentiment scores over the detected SPRTs, whose online news articles are containing noun, the mean, variance, and \( |\text{skewness}| \) for \text{topicsentiscore}(\text{topic}) are measured over \text{topic} \( \in \text{NOTOUNTOPIC}(\text{noun}) \), and they are added as the sentiment features of noun to F2. Here, the mean value of \text{topicsentiscore}(\text{topic}) is equal to \text{sentiscore}_{2}(\text{noun})

Consequently, 10 sentiment features of noun are measured as shown in Table 3, and F22 and F23 are particularly measured to represent the distributions of the sentiment scores of noun over its news and topic.

**Table 3.** Sentiment features of the detected SPRTs’ key noun terms, proposed for this study.

| Feature Sub Set | Level | Sentiment Feature \(^{1,2,3}\) |
|-----------------|-------|---------------------------------|
| F21             | -     | \text{featuresenticore}(\text{noun, pos } = \text{verb}) |
|                 |       | \text{featuresenticore}(\text{noun, pos } = \text{adverb}) |
|                 |       | \text{featuresenticore}(\text{noun, pos } = \text{adjective}) |
|                 |       | \text{nounsentiscore}(\text{noun}) |
|                 |       | \text{sentiscore}_{1}(\text{noun}) |
| F22             | news  | \text{variance and } |\text{skewness}|| of \text{newssentiscore}(\text{news}) |
|                 |       | \text{sentiscore}_{2}(\text{noun}) |
| F23             | topic | \text{variance and } |\text{skewness}|| of \text{topicsentiscore}(\text{topic}) |

Notes:  
1. The statistics of \text{newssentiscore}(\text{news}) and \text{topicsentiscore}(\text{topic}) are obtained respectively over \text{news} \( \in \text{NOUNNEWS}(\text{noun}) \) and \text{topic} \( \in \text{NOTOUNTOPIC}(\text{noun}) \).  
2. If \( n(\text{NOUNNEWS}(\text{noun})) \leq 1 \), the variance value of \text{newssentiscore}(\text{news}) is set as 0. If \( n(\text{NOUNNEWS}(\text{noun})) = 2 \), the skewness value of \text{newssentiscore}(\text{news}) is set as 0.  
3. If \( n(\text{NOTOUNTOPIC}(\text{noun})) \leq 1 \), the variance value of \text{topicsentiscore}(\text{topic}) is set as 0. If \( n(\text{NOTOUNTOPIC}(\text{noun})) = 2 \), the skewness value of \text{topicsentiscore}(\text{topic}) is set as 0.
2.3.3. Measure the Complex Network Structural Features of the SPRTs’ Key Noun Terms

A network whose structure is irregular, complex, and dynamically evolving over time is defined as a complex network. The research on complex networks has resulted in the identification of a series of unifying principles and statistical properties that are common to most real networks [60]. For a given plain graph, approaches that are based on the structure-based patterns of complex networks can be grouped into feature-based and proximity-based approaches: feature-based approaches extract graph-centric features, e.g., node degree; proximity-based approaches quantify the closeness of nodes in the graph to identify associations, e.g., PageRank [61]. In particular, feature-based approaches compute various measures that are associated with the nodes, dyads, triads, egonets, communities, and global graph structure. Among these measures, this paper focuses on the nodes and communities because they both correspond to the node perspective.

Network properties characterize an individual node’s position within a complex network. The three most widely investigated concepts for evaluating such network properties are the degree, closeness, and betweenness centralities [57–59]. These are the standard measures of node centrality, which were originally introduced to quantify the importance of an individual in a social network. Given an adjacency matrix $M_{n \times n} (m_{ij})$ of networks, where $n \geq 3$, the three normalized network centralities can be respectively defined as follows:

$$\text{degree}_i = \frac{1}{(n - 1)} \sum_{j \neq i} m_{ij},$$  \hspace{1cm} (23)

where $m_{ij} = 1$ if node $i$ is connected to node $j$. A high value of degree$_i$ means that node $i$ acts as a center in the network.

$$\text{closeness}_i = (n - 1) \left[ \sum_{j \neq i} d_{ij} \right]^{-1},$$  \hspace{1cm} (24)

where $d_{ij}$ is the number of edges in the shortest path from node $i$ to node $j$. closeness$_i$ indicates the influence of node $i$ on the other nodes.

$$\text{betweenness}_i = \frac{1}{(n - 1)(n - 2) / 2} \sum_{j \neq i \neq k} g_{jk} / g_{jik},$$  \hspace{1cm} (25)

where $g_{jk}$ is the number of the shortest paths between node $j$ and node $k$, and $g_{jik}$ is the number of the shortest paths between node $j$ and node $k$ that contain node $i$. A high betweenness$_i$ value means that node $i$ is located at the core of the networks and has higher momentum of transition.

A community is a densely connected subgroup, which is known to exist in many real-world networks, and community detection (CD) can help us understand networks more deeply and identify interesting properties that are shared by the nodes [62,63]. The fundamental idea behind most CD methods is to partition the nodes of the network into modules [64]. For the agglomerative methods of CD, there are two commonly used algorithms: first, Newman’s CD algorithm is a widely used agglomerative method that uses modularity to measure the goodness of the current partitioning; second, the recently developed Louvain method [65] is an agglomerative method and is commonly used because of its low computational complexity and high performance. When merging communities, the Louvain method considers not only the modularity but also the consolidation ratio [41]. Newman’s algorithm is effective but slow, whereas Louvain’s method is much more computationally efficient [66]. Therefore, this paper adopts the Louvain method for detecting topical communities from complex networks of key noun terms, which are used to label the detected SPRTs.

Based on the abovementioned definitions related to complex networks, this component extracts the complex network structural features regarding noun, namely F3, by constructing two types of the complex networks of the SPRTs’ key noun terms: cross-boundary networks and in-boundary networks. Figure 2 describes how the networks of the key noun terms are constructed respectively, and details are explained as follows:
The cross-boundary networks (CBNs) are constructed by using the key noun terms as nodes, and setting edges by the co-occurrence relationship between the key noun terms in terms of news and topics. In other words, CBN\textsuperscript{co-news} is constructed by making the key noun terms as nodes and their co-occurrence frequencies in online news articles, i.e., co-news frequencies, as the corresponding link weights. Similarly, by setting co-occurrence frequencies in the detected topics, i.e., co-topic frequencies, as the corresponding link weights, CBN\textsuperscript{co-topic} is constructed.

In-boundary networks (IBNs) are built up by using the key noun terms in a particular boundary, and their co-occurrence relationships with respect to the boundary. For IBNs, this study uses two types of boundaries, topics, and communities. First, let ITN\textsuperscript{co-news}(topic) be a kind of IBNs, constructed by setting topic as the boundary and co-news frequencies of the key noun terms as link weights. Second, the Louvain method-based CD on CBN\textsuperscript{co-topic} is performed to take into account the semantic relationship among the key noun terms in terms of their co-topic frequencies. Unlike the TD, the CD allows noun to only one of the detected communities. For each detected community, community, an in-community network, i.e., ICN\textsuperscript{co-topic}(community), is formed by setting co-topic frequencies of the key noun terms in the boundary of community as the link weights.

To evaluate the network properties of noun in both CBN\textsuperscript{co-news} and CBN\textsuperscript{co-topic}, degree, closeness, and betweenness are respectively measured as the complex network structural features.
of noun. Relating to the IBNs, the network properties of noun in ITN\textsuperscript{co-news\_topic} are degree(noun, ITN\textsuperscript{co-news\_topic}), closeness(noun, ITN\textsuperscript{co-news\_topic}), and betweenness(noun, ITN\textsuperscript{co-news\_topic}). In particular, to represent the distribution of the three network centralities of noun over the detected SPRTs, the mean, variance, and |skewness| are measured regarding noun, and they are added as the complex network structural features of noun to F3. Then, the structural properties of noun in its corresponding ICN\textsuperscript{co-topic\_community} are obtained as degree(noun, ICN\textsuperscript{co-topic\_community}), closeness(noun, ICN\textsuperscript{co-topic\_community}), and betweenness(noun, ICN\textsuperscript{co-topic\_community}). As a result, Table 4 shows that 18 complex network structural features of noun for each of the constructed complex networks of the SPRTs’ key noun terms.

Table 4. Complex network structural features of the SPRTs’ key noun terms, proposed for this study.

| Feature Sub Set | Network Type | Boundary Type | Link Type | Complex Network Structure Feature \(^ {1,2}\) |
|-----------------|--------------|---------------|-----------|---------------------------------------------|
| F31             | Cross-boundary | -             | co-news   | degree(noun, CBN\textsuperscript{co-news\_topic}) |
|                 |              |               |           | closeness(noun, CBN\textsuperscript{co-news\_topic}) |
|                 |              |               |           | betweenness(noun, CBN\textsuperscript{co-news\_topic}) |
| F32             | -             | co-topic      |           | degree(noun, ICN\textsuperscript{co-topic\_community}) |
|                 |              |               |           | closeness(noun, ICN\textsuperscript{co-topic\_community}) |
|                 |              |               |           | betweenness(noun, ICN\textsuperscript{co-topic\_community}) |
| F33             | In-boundary   | Given topic   | co-news   | mean, variance, and |skewness| of degree(noun, ITN\textsuperscript{co-news\_topic}) |
|                 |              |               |           | mean, variance, and |skewness| of closeness(noun, ITN\textsuperscript{co-news\_topic}) |
|                 |              |               |           | mean, variance, and |skewness| of betweenness(noun, ITN\textsuperscript{co-news\_topic}) |
| F34             | Given community| co-topic |           | degree(noun, ICN\textsuperscript{co-topic\_community}) |
|                 |              |               |           | closeness(noun, ICN\textsuperscript{co-topic\_community}) |
|                 |              |               |           | betweenness(noun, ICN\textsuperscript{co-topic\_community}) |

Notes: 1 The statistics of degree, closeness, and betweenness of noun in ITN\textsuperscript{co-news\_topic} are obtained over topic \(\in\) NOUNTOPIC(noun). 2 If \(n(\text{NOUNTOPIC(noun)}) \leq 1\), the variance values of degree, closeness, and betweenness of noun in ITN\textsuperscript{co-news\_topic} are set as 0. If \(n(\text{NOUNTOPIC(noun)}) \leq 2\), the skewness values of degree, closeness, and betweenness of noun in ITN\textsuperscript{co-news\_topic} are set as 0.

2.4. Classify the Key Noun Terms of the SPRTs into the SocialTERMs and the EventTERMs

This subsection defines a target variable for classification, and introduces machine learning techniques used for classification in the previous text mining applications. In addition, it explains the experimental settings to generate configurations, which result from combining the different feature sets and different classification techniques.

2.4.1. Definition for a Target Variable

By referring to the examples, mentioned in the introduction, SocialTERM and EventTERM can be defined as below:

**Definition 1.** (SocialTERM) Given social-problem-related topics (SPRTs) and their key noun terms, the SocialTERM of a SPRT is defined as a key noun term that are perceived as: characterizing the SPRT as a social problem; and being a useful cue to identifying and monitoring the ongoing and future events of the social problem. SocialTERMs are irrelevant of the event-specific characteristics of the SPRTs, e.g., when and where the events of the SPRT happened, but reflective of the social-problem-specific perspectives of the SPRTs, e.g., what social problems the SPRT includes, and what causes are underlying such social problems.

**Definition 2.** (EventTERM) Given SPRTs and their key noun terms, the EventTERM of a SPRT is defined as a key noun term that is not perceived as a SocialTERM, because it is not able to explain the social-problem-specific
characteristic of the SPRT but the event-specific characteristics of the events that belong to the SPRT. Thus, the EventTERMs are considered not useful to identifying and monitoring the ongoing and future events of social problems.

For the key noun terms obtained from the detected topics, their target variables, \( y(noun) \), are manually identified by three professional and experienced social scientists, invited as inspectors. Defined as Equation (26), these are used as the true values to be compared to the estimated values.

\[
y(noun) = \begin{cases} 
    \text{SocialTERM} & \text{if } noun \text{ is the social problem-specific key noun term of the detected SPRTs}, \\
    \text{EventTERM} & \text{otherwise}.
\end{cases}
\]

To assure the reliability of the manual investigation, Cohen’s Kappa, \( k \), is calculated for the inter-agreement between the three inspectors, and it is defined as

\[
k = 1 - \left( \frac{1 - p_o}{1 - p_e} \right),
\]

where \( p_o \) is the relative observed agreement among the three inspectors, and \( p_e \) is the hypothetical probability of chance agreement. The Cohen’s Kappa is a statistic that measures inter-rater agreement for categorical items, and it serves as an evidence that the combination of several sources reduced the bias of individual sources \([56,67,68]\). For these reasons, it is adopted in this study to evaluate the consistency of annotated results by the three inspectors.

### 2.4.2. Machine Learning Techniques for Classification in the Previous Text Mining Applications

To distinguish between the SocialTERMs and the EventTERMs among the key noun terms of the detected SPRTs, this paper adopts supervised classification techniques, which have been extensively studied due to their high classification performance. Of the classification techniques that were used in the previous works in Table 1, four commonly used classification techniques and a recently proposed deep-learning-based technique are adopted as base learners for this study. To name, they are C4.5 as Decision Tree (DT) \([9,69]\), Naïve Bayes (NB) \([70–72]\), Radial Basis Function Network (RBFN) \([9]\), Support Vector Machine (SVM) \([73,74]\), and Deep Belief Network (DBN) \([75–77]\). Each of them is explained in the S.1 of Supplementary Materials.

In addition to the five base learners, three types of ensemble methods are combined with each of the five base learners for this study. Ensemble learning is a machine learning paradigm in which multiple learners are trained to solve the same problem. In contrast to the base learners, which try to learn one hypothesis from the training data, the ensemble learning methods try to learn a set of hypotheses and combine them for use. In general, ensemble methods are divided into two categories: instance partitioning and feature partitioning. Bagging and Boosting are instance partitioning methods; RS is a feature partitioning method \([78]\).

Particularly, the three ensemble methods, namely Bagging, Boosting, and RS, are summarized as follows: Bagging is one of the simplest ensemble methods but has surprisingly good performance. The combination strategy of base learners for Bagging is majority voting. This strategy reduces the variance when combined with the base learner generation strategies. Bagging is particularly appealing when the available data are of limited size \([79]\). Unlike Bagging, Boosting produces different base learners by sequentially giving instances that have been misclassified by the previous base learner larger weight in the next iteration of training. The final model that is obtained by Boosting is a linear combination of several base learners, which are weighted by their own performances. There are several Boosting algorithms; the most widely used is AdaBoost \([78]\). RS is an ensemble construction technique, which uses random subspaces to both construct and aggregate the base learners. If a dataset has many redundant or irrelevant features, base learners in random subspaces may be better than in the original
feature space. The combined decision of such base learners may be superior to that of a single classifier that is constructed on the original training dataset in the complete feature sets.

To the best of our knowledge from Table 1, no previous study has compared the performances of the state-of-the-art classification techniques, particularly DBN, in distinguishing between the SocialTERMs and the EventTERMs among the key terms of the detected SPRTs. Hence, this study adopts the five base learners and their combinations with the three ensemble methods. Moreover, these classification techniques are compared in terms of their performances.

2.4.3. Experimental Settings on Features and Classification Techniques

In this paper, the experiments are performed with 60 configurations, which result from combining the three feature sets, namely F1, F1 + F2, and F1 + F2 + F3, and 20 classification techniques. Details on the experimental settings are as follows.

The three types of features, i.e., F1, F2, and F3, are obtained after the feature extraction of the Section 2.3. Based on these different types of features, three feature sets are constructed in an incremental way: feature set F1; feature set F1 + F2; and feature set F1 + F2 + F3. This incremental order implies the evolutionary sequence of features [19,80].

In addition, three popular ensemble methods, i.e., Bagging, Boosting, and RS, are implemented respectively with the five base learners. Consequently, the paper uses 20 classification techniques to differentiate the SocialTERMs from the EventTERMs as described in Table 5. For an experiment that uses one of the 20 classification techniques, a 10 fold validation is performed to train a classifier and evaluate it. Before performing the experiments, if the sample sizes of two classes in \( y(\text{noun}) \) of the data set for an experiment are imbalanced, the imbalanced problem has to be resolved because imbalanced datasets may have problems such as small sample size, overlapping or class separability, and small disjunctions [81]. Previous approaches for dealing with imbalanced datasets are grouped into four categories: algorithm-level, e.g., Hellinger Distance Decision Trees; data-level, e.g., random oversampling and synthetic minority oversampling technique (SMOTE); cost-sensitive, e.g., AdaCost; and classifier ensembles, e.g., Bagging [82]. Among them, the SMOTE approach is known for its good performances when adopted with ensemble methods [81], and therefore it is used to deal with the imbalance problem of this study [83].

Among the 20 classification techniques, to implement the conventional 16 classification approaches of DT, NB, RBFN, and SVM, the data mining toolkit WEKA (Waikato Environment for Knowledge Analysis) version 3.7.0 is used because it is the best-known open-source toolkit with a collection of various machine learning algorithms for solving data mining problems [19,78]. In detail, for the base learners, J48 module (WEKA’s own version of C4.5) for DT, RBFNetwork module for RBFN, NaïveBayes module for NB, and SMO module for SVM; for the ensemble methods, Bagging module for Bagging, AdaBoostM1 module for Boosting, and RandomSubSpace module for RS. Moreover, for DBN and its ensemble learning methods, python-based deep learning tutorials from ‘www.deeplearning.net’ are used as references, and modified. In implementing DBN, the number of hidden layers is set to two, and the dimension in each layer is set to 100 by default.
Table 5. Classification techniques, constructed for this study.

| Classification Techniques | Base Learners | Ensemble Methods |
|--------------------------|---------------|------------------|
|                          | DT | NB | RBFN | SVM | DBN | BL | Bagging | Boosting | RS |
| BL DT                    | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Bagging DT               | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Boosting DT              | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| RS DT                    | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| BL NB                    | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Bagging NB               | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Boosting NB              | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| RS NB                    | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| BL RBFN                  | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Bagging RBFN             | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Boosting RBFN            | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| RS RBFN                  | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| BL SVM                   | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Bagging SVM              | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Boosting SVM             | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| RS SVM                   | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| BL DBN                   | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Bagging DBN              | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| Boosting DBN             | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |
| RS DBN                   | ✓  | ✓  | ✓    | ✓   | ✓   | ✓  | ✓      | ✓       | ✓  |

Notes: 1. BL is the abbreviation of baseline, which means that the ensemble method is not used.

2.5. Evaluate Results with Comparisons

This component assesses the performance of the configurations of three feature sets and 20 classification techniques for classifying the key noun terms of the SPRTs into the SocialTERMs and the EventTERMs. Among the standard metrics, widely used in IR and text classification studies, this paper uses the three performance measures, i.e., accuracy, F-measure, and AUC to evaluate each configuration. In particular, the definition of accuracy can be explained with a confusion matrix as shown in Table 6, and it is defined as

\[
\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN},
\]

and F-measure is obtained by

\[
F - \text{measure} = \frac{2TP}{2TP + FP + FN}.
\]

Table 6. Confusion matrix for classification results.

| Actual Result | SocialTERMs | EventTERMs |
|---------------|-------------|------------|
| Predicted result | SocialTERMs | True positive (TP) | False positive (FP) |
|                | EventTERMs | False negative (FN) | True negative (TN) |
investigated by conducting 60 individual pairwise $t$ tests, i.e., $60 = 3 \times 20$ classification techniques. Moreover, classification techniques for a certain feature set are compared in terms of the three performance measures by conducting 120 individual pairwise $t$ tests, which are composed as follows: 30 between five BL classification techniques, i.e., $30 = 10 \times 3$; 45 between five BL classification techniques and 15 ensemble learning methods, i.e., $45 = 15 \times 3$; and 45 between 15 ensemble learning methods, i.e., $45 = 15 \times 3$.

3. Results

3.1. Test Bed for Data Collection: South Korea and Korean News Portal Site

Relating to the Section 2.1, this paper selected South Korea as a test-bed country for three main reasons: first, it is an information and communication technology (ICT)-intensive nation, so many online news articles are available, and it is easier to identify the SPRTs from online news articles [8,9]. Second, it is well known for its high prevalence of social problems, e.g., it has the highest rate of suicide among OECD countries [84]. This means that South Korea needs to identify social problems more than the other countries do, which corresponds to the desired application of this study. Third, it is a knowledge-intensive country, so, once identified, SocialTERMs can be better used to explore technologies for solving social problems than in other countries [85].

By using the distributed web-crawling program, the online news articles were collected from NAVER.com, which is the best-known Korean news portal site. These articles had been published in the society-related news sections in the 356 days from May 2013 to June 2014, i.e., $t = 1, \ldots, 365$. In total, 126,402 online news articles were collected from the targeted society-related sections, and the parsed data were stored in the relational database for the experiments.

3.2. Evaluation Results

Relating to the Section 2.2, 43,711 online news articles with negative sentiment were selected from the collected 126,403 online news articles. Next, the thresholds $\varepsilon = 0.3$, $\alpha = 20$, and $\beta = 10$ were determined based on a pre-topic analysis of 100 online news articles, which were published in the first month, and 2961 topics of online news articles, which were detected from the 43,711 online news articles by Algorithm 1. Among the 2961 detected topics of online news articles, the 467 topics with more than 10 ($= \beta$) online news articles were chosen as the final detected topics, namely the SPRTs. Then, as explained in the Section 2.3, the three types of features, namely temporal weight, sentiment, and complex network structural features, were measured for the 1810 key noun terms, which were extracted from the 467 detected SPRTs (see Table 7 for examples of the 1810 key noun terms). Particularly in measuring the complex network structural features, JUNG (http://jung.sourceforge.net/), which is a Java-based software library for network analysis, was used to obtain the network centralities, and Gephi (https://gephi.org/), which is an open-source graph visualization platform, was used to identify communities from the constructed co-news and co-topic key term networks. Tables A3–A6 in Appendix B show the descriptive statistics on the three types of features that represent the 1810 key noun terms.
Table 7. The list of the top 10 key noun terms for each class according to their document frequencies, dfscore\(_1\) (noun).

| Class     | Key Noun Terms in Korean (English) | dfscore\(_1\) (noun) | In-Class Rank | Overall Rank |
|-----------|-----------------------------------|----------------------|---------------|--------------|
| SocialTERM | 여성 (female)                      | 0.0611               | 1             | 5            |
|           | 대학 (university)                 | 0.0611               | 2             | 6            |
|           | 병원 (hospital)                   | 0.0604               | 3             | 8            |
|           | 아아 (child)                      | 0.0589               | 4             | 10           |
|           | 학생 (student)                    | 0.0573               | 5             | 11           |
|           | 환자 (patient)                    | 0.0565               | 6             | 12           |
|           | 장애 (disability)                 | 0.0553               | 7             | 13           |
|           | 남성 (male)                       | 0.0528               | 8             | 21           |
|           | 선고 (sentence)                   | 0.0515               | 9             | 23           |
| EventTERM | 경찰 (prosecution)                | 0.0511               | 10            | 24           |
|           | 경찰 (police)                     | 0.0776               | 1             | 1            |
|           | 학교 (school)                     | 0.0645               | 2             | 2            |
|           | 부산 (Busan)                      | 0.0642               | 3             | 3            |
|           | 교육 (education)                  | 0.0614               | 4             | 4            |
|           | 사람 (person)                     | 0.0609               | 5             | 7            |
|           | 대구 (Daegu)                      | 0.0603               | 6             | 9            |
|           | 수사 (investigation)              | 0.0550               | 7             | 14           |
|           | 회장 (president)                  | 0.0543               | 8             | 15           |
|           | 발생 (outbreak)                   | 0.0543               | 9             | 16           |
|           | 교수 (professor)                  | 0.0543               | 10            | 17           |

The target variables, which are denoted as \(y\) (noun), of the 1810 key noun terms were manually identified by three inspectors. The procedure yielded a Cohen’s Kappa inter-rater reliability of 0.8678, thereby indicating good agreement, i.e., \(k \geq 0.8\), according to Lombard et al. [86]. Disagreements among the three inspectors were jointly reviewed until a final agreement was reached. These were used as the true values, to be compared to the estimated values. In addition, the 1810 key noun terms were the imbalanced in terms of the classes of their target variables. To resolve this imbalance issue, SMOTE was applied to the 1810 key noun terms. By adding 502 new instances of \(y\) (noun) = SocialTERM, a balanced data set of 1156 SocialTERMs and 1156 EventTERMs was prepared for the following experiments.

Next, according to the Section 2.4, experiments were performed on the prepared data set, and Table 8 shows the experimental results on the three performance measures for different feature sets and different classification techniques. Consequently, the full feature set configuration of \(F_1 + F_2 + F_3\) and the ensemble learning method, namely Boosting DT, gave the best accuracy, i.e., 83.8769%, which is 1.3264% better than the second best configuration, i.e., \(F_1 + F_2\) and Boosting DT. Moreover, Table 8 shows that, with \(F_1 + F_2 + F_3\), Boosting DT also gave the best performances in terms of \(F\)-measure (1.7112% better than with \(F_1 + F_2\) and ACU (1.8174% better than with \(F_1 + F_2\)). Thus, the results in Table 8 provided an answer to the part of RQ1 that is how well the three types of features perform by using different classification techniques.

The possible reason for the best performances of Boosting DT is as follows: DT could deal with the numerical features of this study properly as categorical features; and DT with Boosting could reduce multi-collinearity problems, which may exist among features [9,74,78].
Table 8. Performances of different feature sets and different classification techniques.

(a) Performance Measure = Accuracy

| Feature set | DT | BL DT | Bagging DT | Boosting DT | RS DT |
|-------------|----|-------|------------|-------------|-------|
| F1          | 58.778 ± 0.0059 | 74.384 ± 0.0076 | 63.2497 ± 0.0158 | 60.6012 ± 0.0071 |
| F1 + F2     | 66.3552 ± 0.0042 | 81.1332 ± 0.0064 | 82.5505 ± 0.0118 | 70.7771 ± 0.0112 |
| F1 + F2 + F3| 66.5297 ± 0.0039 | 81.7142 ± 0.0059 | 83.8769 ± 0.0039 | 75.1081 ± 0.0089 |

| Feature set | NB | BL NB | Bagging NB | Boosting NB | RS NB |
|-------------|----|-------|------------|-------------|-------|
| F1          | 60.2984 ± 0.0035 | 58.9720 ± 0.0022 | 59.6583 ± 0.0040 | 58.7659 ± 0.0022 |
| F1 + F2     | 60.4700 ± 0.0020 | 62.1396 ± 0.0027 | 63.4530 ± 0.0049 | 62.2506 ± 0.0022 |
| F1 + F2 + F3| 60.6142 ± 0.0021 | 62.0401 ± 0.0040 | 64.1955 ± 0.0058 | 62.1799 ± 0.0034 |

| Feature set | RBFN | BL RBFN | Bagging RBFN | Boosting RBFN | RS RBFN |
|-------------|------|---------|--------------|---------------|--------|
| F1          | 58.778 ± 0.0059 | 70.9400 ± 0.0078 | 59.5603 ± 0.0057 | 72.4308 ± 0.0062 |
| F1 + F2     | 66.3552 ± 0.0042 | 76.9867 ± 0.0058 | 65.8276 ± 0.0061 | 78.4386 ± 0.0061 |
| F1 + F2 + F3| 66.5297 ± 0.0039 | 77.1597 ± 0.0067 | 68.9821 ± 0.0034 | 78.5092 ± 0.0056 |

| Feature set | SVM | BL SVM | Bagging SVM | Boosting SVM | RS SVM |
|-------------|-----|--------|-------------|--------------|-------|
| F1          | 59.1825 ± 0.0026 | 59.4132 ± 0.0037 | 59.3267 ± 0.0031 | 59.3426 ± 0.0031 |
| F1 + F2     | 63.0926 ± 0.0024 | 64.4464 ± 0.0041 | 64.6958 ± 0.0037 | 64.5732 ± 0.0028 |
| F1 + F2 + F3| 63.2930 ± 0.0028 | 65.7434 ± 0.0041 | 65.3720 ± 0.0032 | 65.3777 ± 0.0031 |

| Feature set | DBN | BL DBN | Bagging DBN | Boosting DBN | RS DBN |
|-------------|-----|--------|-------------|--------------|-------|
| F1          | 53.3990 ± 0.0170 | 54.6916 ± 0.0178 | 57.5148 ± 0.0115 | 49.6817 ± 0.0085 |
| F1 + F2     | 60.1239 ± 0.0174 | 61.5670 ± 0.0115 | 61.6693 ± 0.0144 | 59.9632 ± 0.0183 |
| F1 + F2 + F3| 60.7555 ± 0.0152 | 62.6925 ± 0.0095 | 61.7397 ± 0.0136 | 60.6326 ± 0.0153 |

(b) Performance measure = F-measure

| Feature set | DT | BL DT | Bagging DT | Boosting DT | RS DT |
|-------------|----|-------|------------|-------------|-------|
| F1          | 57.6430 ± 0.0082 | 74.3585 ± 0.0088 | 63.6145 ± 0.0159 | 59.3495 ± 0.0111 |
| F1 + F2     | 65.4532 ± 0.0039 | 81.1585 ± 0.0052 | 82.1295 ± 0.0119 | 70.5122 ± 0.0110 |
| F1 + F2 + F3| 65.5590 ± 0.0048 | 81.7649 ± 0.0072 | 83.8407 ± 0.0068 | 75.1471 ± 0.0113 |

| Feature set | NB | BL NB | Bagging NB | Boosting NB | RS NB |
|-------------|----|-------|------------|-------------|-------|
| F1          | 60.4035 ± 0.0030 | 58.8007 ± 0.0028 | 59.5686 ± 0.0043 | 58.6722 ± 0.0024 |
| F1 + F2     | 56.8606 ± 0.0023 | 62.1944 ± 0.0033 | 63.7492 ± 0.0050 | 62.2347 ± 0.0019 |
| F1 + F2 + F3| 57.0204 ± 0.0027 | 61.7321 ± 0.0029 | 64.3634 ± 0.0065 | 61.8825 ± 0.0028 |

| Feature set | RBFN | BL RBFN | Bagging RBFN | Boosting RBFN | RS RBFN |
|-------------|------|---------|--------------|---------------|--------|
| F1          | 57.6430 ± 0.0082 | 70.6085 ± 0.0068 | 58.1797 ± 0.0057 | 72.1917 ± 0.0083 |
| F1 + F2     | 65.4532 ± 0.0039 | 76.9010 ± 0.0066 | 64.8197 ± 0.0052 | 78.2892 ± 0.0075 |
| F1 + F2 + F3| 65.5590 ± 0.0048 | 77.2176 ± 0.0068 | 68.8268 ± 0.0044 | 78.5373 ± 0.0057 |

| Feature set | SVM | BL SVM | Bagging SVM | Boosting SVM | RS SVM |
|-------------|-----|--------|-------------|--------------|-------|
| F1          | 58.4588 ± 0.0028 | 58.7076 ± 0.0045 | 58.5790 ± 0.0025 | 58.5808 ± 0.0023 |
| F1 + F2     | 61.5544 ± 0.0019 | 63.9253 ± 0.0038 | 64.0992 ± 0.0034 | 63.8106 ± 0.0027 |
| F1 + F2 + F3| 61.7701 ± 0.0027 | 65.0890 ± 0.0039 | 65.0543 ± 0.0033 | 64.9956 ± 0.0030 |

| Feature set | DBN | BL DBN | Bagging DBN | Boosting DBN | RS DBN |
|-------------|-----|--------|-------------|--------------|-------|
| F1          | 43.4786 ± 0.0401 | 47.3221 ± 0.0377 | 52.3530 ± 0.0311 | 33.3678 ± 0.0116 |
| F1 + F2     | 53.9359 ± 0.0327 | 56.9606 ± 0.0181 | 56.6800 ± 0.0266 | 53.5839 ± 0.0340 |
| F1 + F2 + F3| 54.8043 ± 0.0290 | 58.7171 ± 0.0140 | 56.8986 ± 0.0231 | 54.4863 ± 0.0305 |
Table 8. Cont.

| Feature set | DT | BL DT | Bagging DT | Boosting DT | RS DT |
|-------------|----|-------|------------|-------------|-------|
| F1          | 61.3058 ± 0.0072 | 82.3743 ± 0.0064 | 70.3160 ± 0.0180 | 65.8481 ± 0.0096 |
| F1 + F2     | 68.7706 ± 0.0082 | 88.0634 ± 0.0046 | 89.8433 ± 0.0112 | 77.0963 ± 0.0106 |
| F1 + F2 + F3| 69.5211 ± 0.0095 | 88.6718 ± 0.0049 | 91.6607 ± 0.0054 | 79.8530 ± 0.0077 |

| NB          | BL NB | Bagging NB | Boosting NB | RS NB |
|-------------|-------|------------|-------------|-------|
| F1          | 64.1659 ± 0.0018 | 62.8991 ± 0.0015 | 62.9606 ± 0.0038 | 62.9253 ± 0.0016 |
| F1 + F2     | 69.7733 ± 0.0037 | 67.4650 ± 0.0015 | 69.0069 ± 0.0032 | 67.2505 ± 0.0019 |
| F1 + F2 + F3| 69.8820 ± 0.0029 | 67.3182 ± 0.0023 | 70.0062 ± 0.0043 | 66.9227 ± 0.0023 |

| RBFN        | BL RBFN | Bagging RBFN | Boosting RBFN | RS RBFN |
|-------------|---------|--------------|---------------|---------|
| F1          | 61.3058 ± 0.0072 | 77.6983 ± 0.0061 | 63.1809 ± 0.0053 | 79.3792 ± 0.0068 |
| F1 + F2     | 68.7706 ± 0.0082 | 84.4544 ± 0.0049 | 72.4777 ± 0.0039 | 85.5619 ± 0.0047 |
| F1 + F2 + F3| 69.5211 ± 0.0095 | 84.9973 ± 0.0052 | 74.3151 ± 0.0033 | 85.9469 ± 0.0052 |

| SVM         | BL SVM | Bagging SVM | Boosting SVM | RS SVM |
|-------------|--------|------------|--------------|-------|
| F1          | 59.1926 ± 0.0029 | 61.7137 ± 0.0038 | 62.3229 ± 0.0050 | 59.3262 ± 0.0022 |
| F1 + F2     | 63.2656 ± 0.0018 | 68.0528 ± 0.0035 | 68.9031 ± 0.0035 | 64.5211 ± 0.0027 |
| F1 + F2 + F3| 63.4443 ± 0.0021 | 69.4318 ± 0.0036 | 69.9604 ± 0.0044 | 65.3753 ± 0.0029 |

| DBN         | BL DBN | Bagging DBN | Boosting DBN | RS DBN |
|-------------|--------|------------|--------------|-------|
| F1          | 53.7798 ± 0.0147 | 55.3225 ± 0.0140 | 56.6244 ± 0.0130 | 50.2384 ± 0.0045 |
| F1 + F2     | 60.2221 ± 0.0170 | 61.9084 ± 0.0095 | 60.9947 ± 0.0156 | 60.0628 ± 0.0179 |
| F1 + F2 + F3| 60.8064 ± 0.0147 | 62.9222 ± 0.0084 | 61.2206 ± 0.0147 | 60.6341 ± 0.0164 |

Notes: ± are standard deviations. For each base learner, the best result is highlighted as italics, and the best result over all configurations is additionally highlighted as red and bold italics.

4. Discussion

4.1. Comparisons of Feature Sets

Table A6 of Appendix C shows the comparison results of pairwise t tests, which were performed to evaluate the effects of different feature sets on the performance of a classification technique in terms of the three performance measures. The comparison results gave answers to the part of RQ1 about which feature set and features give the best results, and their details are as follows.

By summarizing the comparison results in Table A6, Figure 3 illustrates the ratio of agreement with the positive effect of adding a feature subset on increasing performance from different perspectives. One of its key findings is that for most of the 20 classification techniques, adding F1, F2, and F3 individually increased performance in respect of the three performance measures. This indicates that each of the feature sets that were suggested by this study is useful for identifying the SocialTERMs from the detected SPRTs from online news articles. Sentiment feature set F2 led to better performances, regardless of the classification technique. The effect of adding complex network structural feature set F3 was smaller than those of adding F1 and F2.
Figure 3. Ratio of agreement that a feature set improved performance.

Furthermore, using Boosting DT, which was shown to be the best classification technique in Table 8, this paper performed pairwise t tests to compare their different feature subsets, and investigated the effect of adding each feature subset on the classification performance. Table 9 shows that for all three performance measures, the significant performance improvements by feature sets F1, F2, and F3 was respectively attributed to feature subsets F11 and F12 for F1, F21 and F23 for F2, and F31 for F3. This indicates that these features are more useful in characterizing the relatedness of the key noun terms to social problems.
Table 9. Pairwise t tests on three performance measures for different feature subsets when the best classification technique, namely Boosting DT, was selected.

(a) Performance Measure = Accuracy

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F1          | F11                   | F1 (+) + F11 > F1 (+) | 4.7454 0.0000 | ✓         |
| F12         | F1 (+) + F11 + F12 > F1 (+) + F11 | 4.0312 0.0002 | ✓         |
| F13         | F1 (+) + F11 + F12 + F13 > F1 (+) + F11 + F12 | −1.3335 0.1877 |
| F14         | F1 (+) + F11 + F12 + F13 + F14 > F1 (+) + F11 + F12 + F13 | 0.0988 0.9216 |

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F2          | F21                   | F2 (+) + F21 > F2 (+) | 21.4020 0.0000 | ✓         |
| F22         | F2 (+) + F21 + F22 > F2 (+) + F21 | −1.1473 0.2560 |
| F23         | F2 (+) + F21 + F22 + F23 > F2 (+) + F21 + F22 | 2.2101 0.0312 |

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F3          | F31                   | F3 (+) + F31 > F3 (+) | 3.2401 0.0020 | ✓         |
| F32         | F3 (+) + F31 + F32 > F3 (+) + F31 | 0.8737 0.3862 |
| F33         | F3 (+) + F31 + F32 + F33 > F3 (+) + F31 + F32 | 1.8535 0.0689 |
| F34         | F3 (+) + F31 + F32 + F33 + F34 > F3 (+) + F31 + F32 + F33 | 0.5195 0.6054 |

(b) Performance measure = F-measure

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F1          | F11                   | F1 (+) + F11 > F1 (+) | 3.8147 0.0004 | ✓         |
| F12         | F1 (+) + F11 + F12 > F1 (+) + F11 | 5.7016 0.0000 | ✓         |
| F13         | F1 (+) + F11 + F12 + F13 > F1 (+) + F11 + F12 | −0.7661 0.4467 |
| F14         | F1 (+) + F11 + F12 + F13 + F14 > F1 (+) + F11 + F12 + F13 | 1.3614 0.1787 |

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F2          | F21                   | F2 (+) + F21 > F2 (+) | 27.3757 0.0000 | ✓         |
| F22         | F2 (+) + F21 + F22 > F2 (+) + F21 | −0.5322 0.5966 |
| F23         | F2 (+) + F21 + F22 + F23 > F2 (+) + F21 + F22 | 2.0480 0.0451 |

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F3          | F31                   | F3 (+) + F31 > F3 (+) | 2.2752 0.0266 | ✓         |
| F32         | F3 (+) + F31 + F32 > F3 (+) + F31 | 1.6143 0.1128 |
| F33         | F3 (+) + F31 + F32 + F33 > F3 (+) + F31 + F32 | 0.7423 0.4610 |
| F34         | F3 (+) + F31 + F32 + F33 + F34 > F3 (+) + F31 + F32 + F33 | 1.4505 0.1523 |

(c) Performance measure = AUC

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F1          | F11                   | F1 (+) + F11 > F1 (+) | 4.3216 0.0001 | ✓         |
| F12         | F1 (+) + F11 + F12 > F1 (+) + F11 | 6.8192 0.0000 | ✓         |
| F13         | F1 (+) + F11 + F12 + F13 > F1 (+) + F11 + F12 | 0.8119 0.4202 |
| F14         | F1 (+) + F11 + F12 + F13 + F14 > F1 (+) + F11 + F12 + F13 | 0.5489 0.5853 |

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F2          | F21                   | F2 (+) + F21 > F2 (+) | 20.2631 0.0000 | ✓         |
| F22         | F2 (+) + F21 + F22 > F2 (+) + F21 | −1.0577 0.3038 |
| F23         | F2 (+) + F21 + F22 + F23 > F2 (+) + F21 + F22 | 2.2935 0.0256 |

| Feature set | Added feature sub set | Hypothesis | Boosting DT | Supported |
|-------------|-----------------------|------------|-------------|-----------|
| F3          | F31                   | F3 (+) + F31 > F3 (+) | 3.0933 0.0031 | ✓         |
| F32         | F3 (+) + F31 + F32 > F3 (+) + F31 | 2.7597 0.0082 | ✓         |
| F33         | F3 (+) + F31 + F32 + F33 > F3 (+) + F31 + F32 | 1.8379 0.0714 |
| F34         | F3 (+) + F31 + F32 + F33 + F34 > F3 (+) + F31 + F32 + F33 | 1.1500 0.2550 |

Notes: F1 (+) = F2 + F3. F2 (+) = F1 + F3. F3 (+) = F1 + F2. The results are t and p values of the t tests for feature set comparisons, and the results with a significance level higher than 5% are italicized.

4.2. Comparisons on Classification Techniques

In addition, the classification techniques were compared in three ways: base learner vs. base learner (see Table A7 of Appendix C), base learner vs. ensemble learning method (see Table A8 of Appendix C), and ensemble method vs. ensemble method (see Table A9 of Appendix C). Table A7 shows the results of the pairwise t tests, which were performed to examine the effects of different base learners on three performance measures for a specific feature set, and Figure 4 provides an overview of the results in Table A7. According to the results, the performance rankings of all five base learners are...
different according to the selected feature sets, and it implies that there is no single best classification technique for all three performance measures.

![Figure 4. The performance rankings of base learners for a given feature set.](image)

Table A8 shows the results of the pairwise $t$ tests, which were performed to examine the effect of combining an ensemble method on three performance measures for a specific feature set, and Figure 5 summarizes the results in Table A8. To explain, Figure 5 shows that in terms of all three performance measures, combining Bagging yielded better performances than their base learners in most configurations for all the incremental feature sets, while Boosting and RS did not perform as well as Bagging. The reason for the positive effect of Bagging can be that Bagging helps preserve the important information better than the base learners by considering the features in their entirety, unlike the base learners, which only considers the average of the aggregated features. Overall, it is concluded that combining an ensemble learning method is appropriate for this study to identify the SocialTERMs from the detected SPRTs.
unlike the base learners, which only considers the average of the aggregated features. Overall, it is concluded that combining an ensemble learning method is appropriate for this study to identify the SocialTERMs from the detected SPRTs.

Table A9 shows the results of the pairwise $t$ tests, which were performed to examine the effects of different ensemble methods on three performance measures for a specific feature set if a base learner is given. Figures 6–8 explain the performance rankings of the ensemble methods, which are evaluated based on the results in Table A9.

Some interesting findings from Figure 6 are as follows: While Bagging was ranked best among the three ensemble methods if combined with DT and DBN for $F_1$, Boosting gave better accuracies with DT and DBN for $F_1 + F_2$ and $F_1 + F_2 + F_3$, and with NB for all feature sets. The possible reasons for the superiority of Boosting with DT, NB, and DBN are as follows: The strategy of Boosting, which gives higher weights to misclassifications in training, was effective for training models of DT, NB, and DBN with more features; and Boosting’s robustness against the multi-collinearity problems among complex features could help DT, NB, and DT to have better accuracies. Moreover, for all feature sets, RS always achieved better accuracies than other ensemble methods if it was used with RBFN, while no major ensemble method achieved better accuracy with SVM. In Figures 7 and 8, the same results with Figure 6 were observed, except that for all feature sets, Boosting gave better AUCs if combined with SVM, followed in descending order of Bagging and RS as shown in Figure 8d.
same results with Figure 6 were observed, except that for all feature sets, Boosting gave better AUCs if combined with SVM, followed in descending order of Bagging and RS as shown in Figure 8d. Thus, Figures 6–8 indicate that the choice of an ensemble method for obtaining better performances depends on the feature sets and the base learners. Therefore, it can hardly be said that a single ensemble method gave the best accuracy for the all feature sets with any single base learner. However, as shown in Figure 6f, if the accuracy rankings of the ensemble methods were averaged over the different base learners for a given feature set, Boosting was a comparatively better choice as an ensemble method for any base learner. Moreover, Figures 7f and 8f that averaged the F-measure rankings and the AUC rankings, respectively, also demonstrate the same results with Figure 6f, i.e., the superiority of Boosting over Bagging and RS.

**Figure 6.** The accuracy rankings of ensemble methods for a given feature set with different base learners.

**Figure 7.** The F-measure rankings of ensemble methods for a given feature set with different base learners.

**Figure 8.** The AUC rankings of ensemble methods for a given feature set with different base learners.

5. Conclusions

This paper proposed and examined an automatic approach, namely SocialTERM-Extractor, for distinguishing between the SocialTERMs and the EventTERMs among the key noun terms of the detected SPRTs from a large number of Korean online news articles. It aimed at resolving the challenging issues that were mentioned in Section 1. Using the best-known news portal site of South
5. Conclusions

This paper proposed and examined an automatic approach, namely SocialTERM-Extractor, for distinguishing between the SocialTERMs and the EventTERMs among the key noun terms of the detected SPRTs from a large number of Korean online news articles. It aimed at resolving the challenging issues that were mentioned in Section 1. Using the best-known news portal site of South Korea as a test-bed, experiments were conducted by following the proposed research framework, as explained in Section 2. The experimental results in Table 8 showed that the configuration of the full feature set, namely F1 + F2 + F3 and Boosting DT gave the best performances for accuracy, as well as F-measure and AUC. Its high performances, e.g., 83.8769% accuracy, implies that the proposed approach can automatically identify the SocialTERMs in a reliable way (RQ1 was partly answered).

Furthermore, according to Figure 3, the pairwise t tests on three performance measures for adding a feature set in Table A6 indicated that most of the 20 classification techniques agreed that the three feature sets, namely F1, F2, and F3, contributed to improving the classification performance in a statistically significant way. In particular, it was agreed by all 20 classification techniques that adding sentiment feature set F2 improved the classification performance, in particular unanimously in terms of accuracy and AUC. When the best classification technique, namely Boosting DT, was used, Table 9 showed that the individual addition of feature subsets such as F11, F12, F21, F23, and F31 increased all three performance measures actually. This indicates that the significant improvement in terms of

Thus, Figures 6–8 indicate that the choice of an ensemble method for obtaining better performances depends on the feature sets and the base learners. Therefore, it can hardly be said that a single ensemble method gave the best accuracy for the all feature sets with any single base learner. However, as shown in Figure 6f, if the accuracy rankings of the ensemble methods were averaged over the different base learners for a given feature set, Boosting was a comparatively better choice as an ensemble method for any base learner. Moreover, Figures 7f and 8f that averaged the F-measure rankings and the AUC rankings, respectively, also demonstrate the same results with Figure 6f, i.e., the superiority of Boosting over Bagging and RS.
three performance measures by adding feature sets in Table A6 is attributed to such feature subsets (RQ1 was partly answered).

Relating to the comparisons of the classification techniques, according to Figure 4 (and Table A7), the performance rankings of all five base learners differed according to the selected feature sets (RQ2 was answered). In addition, Figure 5 (and Table A8) revealed that most of the 20 configurations agreed that most ensemble learning methods produced better performances than the base learners (RQ3 was answered). According to Figures 6–8 (and Table A9), the ensemble method that obtains the best results depends on the feature sets and the base learners. Nevertheless, when the performance rankings of an ensemble method for a feature set were averaged over all types of base learners, ensemble learning methods with Boosting showed comparatively better results for all feature sets (RQ3 was answered).

Theoretically, this paper contributes to expanding the related literature by applying text mining and machine learning techniques to a large number of online news articles as big data. To the best of our knowledge, this study is the first to provide an automatic approach for identifying and predicting the SocialTERMs of the detected SPRTs from online news articles. The appropriate SocialTERMs can be identified automatically so anybody, even someone who is unfamiliar with the ongoing social problems, can benefit from the automatic approach of this study. It helps enable everyone to recognize the landscape of the SPRTs from a large amount of event-related textual data without difficulty. In addition, this study has a significant impact on sustainability, since the SocialTERMs can be used as key noun terms in searching for technologies that are helpful for solving social problems and monitoring the ongoing and future events that are associated with the social problems. Eventually, the paper may facilitate innovations in our society by driving the development of technologies for ongoing and future social problems.

Practically, by answering RQ1–RQ2, this paper provided a reference and guidance for researchers, government officials, politicians, and companies that are in need of the system implementation. The paper investigated which kinds of feature sets are preferable, what kinds of classification techniques perform better, and how these two factors must be combined to obtain the best results. These results help determine the proper model for building a system with real-world large data. In the suggested research framework, the paper suggested novel approaches for representing the key noun terms: temporal weight, sentiment, and complex network structural features. Moreover, the paper compared state-of-the-art techniques, including the recently proposed DBN, which is a deep-learning-based technique. It showed that the simpler conventional classification method was better for this study, while the more complex DBN gave worse results. This indicates that the deep architecture is not a magic key for all kinds of problems in machine learning research, as it is known that the deep architecture works for big data cases with a lot of variables. However, as the results were not much worse compared to the other approaches, better performances by the deep architecture in the other applications may be possible.

Thus, if the automatic approach is implemented by developing a system, the system can automatically recommend the SocialTERMs, which are useful key noun terms for exploring technologies that can be used to solve social problems. The SocialTERMs can be applied to the prediction of future social problems and the monitoring the ongoing social problems from a large number of online news articles. Thus, this study finally helps obtain the new insights about how to identify ongoing and upcoming social problems from big data, thereby paving a way to big-data-driven social and technological innovations for the public good.

Further research can be conducted to overcome the limitations of this study. First, this study used only a large number of online news articles, but, in addition to online news articles, large-scale data from social media, e.g., YouTube, Twitter, and Facebook, may provide good sources for extracting temporal weight, sentiment, and complex network structural features on the key noun terms of the detected SPRTs. Second, the paper focused on the three types of features, but there may be other useful features, and more sophisticated classification techniques can be taken into account to improve the classification performance.
In addition, as future work, a portal site that provides the proposed methodology can be planned so that this methodology can be available to individuals and groups who are in need of identifying the SPRTs and their SocialTERMs. The easier-to-use method can also be considered in developing the portal site, e.g., k-means and latent Dirichlet allocation (LDA) for the TD approach. If developed, the proposed methodology and system can be evaluated in terms of whether they are helpful for users not only in exploring technologies for solving social problems but also in monitoring ongoing and future social problems based on a large amount of event-related textual data.

**Supplementary Materials:** The following are available online at http://www.mdpi.com/2071-1050/11/1/196/s1, S.1: Reviews on the State-of-the-Art Machine Learning Techniques.

**Funding:** This study was supported by the National Research Foundation of Korea Grant (NRF-2017R1C1B1010065), funded by the Korean Government.

**Acknowledgments:** I am grateful to the inspectors, who manually coded the extracted key noun terms of the detected SPRTs into SocialTERMs and EventTERMs. I would like to thank the anonymous reviewers for their valuable comments that helped revise the original version of this paper.

**Conflicts of Interest:** The author declares no conflict of interest.
Appendix A

Table A1. Examples of Korean sentiment features.

| English Sentiment Feature | POS    | Final Sentiment Score | Translation   | Morphological Analysis | Korean Sentiment Feature | POS      | Sentiment Value |
|---------------------------|--------|-----------------------|---------------|------------------------|--------------------------|----------|-----------------|
| soil                      | Verb   | −0.7500               | 디럽히다       | 디럽히/pvg+다/ef       | 디럽히/pvg+다/ef          | Verb     | −0.7500         |
| disheartened              | Adjective | −0.4250              | 속상한       | 속상/ncps+하/ssms+Jon/etm | 속상/ncps+하/ssms+Jon/etm | Adjective | −0.4250         |
| instantly                 | Adjective | −0.3750              |               |                        |                          |          |                 |
| straight_away             | Adverb  | −0.5000               | 즉시         | 즉시/mag               | 즉시/mag                 | Adverb   | −0.2000         |
| right_away                | Adverb  | −0.1875               |               |                        |                          |          |                 |
| at_once                   |         | 0.2500                |               |                        |                          |          |                 |
| swiftly                   |         | 0.2500                |               |                        |                          |          |                 |

Table A2. Korean sentiment features, extended from the stem ‘디럽히/pvg’.

| Stem              | Endings | Extended Korean Sentiment Feature | POS | Tense | Sentiment Value |
|-------------------|---------|----------------------------------|-----|-------|-----------------|
| 디럽히/pvg         | 놀다/ef | 디럽히/pvg+다/ef                   | Verb| Present|                 |
|                   | 놀다/ef | 디럽히/pvg+였다/ef                 |     |       |                 |
|                   | 놀 것이다/ef | 디럽히/pvg+것이다/ef |     |       |                 |
|                   | 놀 는/etm | 디럽히/pvg+는/etm                 |     |       |                 |
|                   | 놀 있던/etm | 디럽히/pvg+있던/etm             |     |       |                 |
|                   | 놀 이/etm | 디럽히/pvg+이/etm                 |     |       |                 |
|                   | 놀 계/ecs | 디럽히/pvg+계/ecs                 |     |       |                 |
|                   | 놀 는다/ef | 디럽히/pvg+지/ecs 앞/px+는다/ef  | Adverb| Present|                 |
|                   | 놀 있있다/ef | 디럽히/pvg+지/ecs 앞/px+있다/ef |     |       |                 |
|                   | 놀 이/etm | 디럽히/pvg+이/etm                 |     |       |                 |
|                   | 놀 계/ecs | 디럽히/pvg+계/ecs                 |     |       |                 |
| 지/ecz 앞/px       | 놀 것이다/ef | 디럽히/pvg+것이다/ecz 앞/px+것이다/etm |     |       |                 |
|                   | 놀 는/etm | 디럽히/pvg+는/ecz 앞/px+는/etm    |     |       |                 |
|                   | 놀 있던/etm | 디럽히/pvg+있던/ecz 앞/px+있던/etm |     |       |                 |
|                   | 놀 이/etm | 디럽히/pvg+이/ecz 앞/px+이/etm    |     |       |                 |
|                   | 놀 계/ecs | 디럽히/pvg+계/ecz 앞/px+계/ecs    |     |       |                 |

Sustainability 2019, 11, 196
## Appendix B

Table A3. Descriptive statistics on the features F1 from the test bed.

| Type   | Features         | SocialTERMs | EventTERMs | SocialTERMS + EventTERMs |
|--------|------------------|-------------|------------|--------------------------|
|        |                  | Min        | Max        | Mean        | SD        | Min        | Max        | Mean        | SD        | Min        | Max        | Mean        | SD        |
| F11    | $dfscore_{1} (\text{noun})$ | 0.0007 | 0.0611 | 0.0171 | 0.0123 | 0.0002 | 0.0776 | 0.0144 | 0.0116 | 0.0002 | 0.0776 | 0.0154 | 0.0119 |
|        |                  | variance    | 0.0000 | 0.0090 | 0.0009 | 0.0008 | 0.0000 | 0.0044 | 0.0007 | 0.0006 | 0.0000 | 0.0090 | 0.0008 |
|        |                  | $|\text{skewness}|$   | 0.0768 | 12.3427 | 2.9182 | 1.8452 | 0.0035 | 14.7584 | 3.2080 | 2.0100 | 0.0035 | 14.7584 | 3.1033 |
| F12    | $tfscore_{1} (\text{noun})$ | 0.0000 | 0.3335 | 0.0614 | 0.0528 | 0.0000 | 0.3345 | 0.0430 | 0.0469 | 0.0000 | 0.3345 | 0.0497 | 0.0499 |
|        |                  | variance    | 0.0000 | 0.1813 | 0.0427 | 0.0310 | 0.0000 | 0.1984 | 0.0308 | 0.0285 | 0.0000 | 0.1984 | 0.0351 |
|        |                  | $|\text{skewness}|$   | 0.8417 | 19.1050 | 4.8521 | 2.6011 | 0.0035 | 14.7584 | 3.2080 | 2.0100 | 0.0035 | 14.7584 | 3.1033 |
| F13    | $titlescore_{1} (\text{noun})$ | 0.0471 | 2.9845 | 1.2800 | 0.8035 | 0.0205 | 3.1060 | 1.1290 | 0.7854 | 0.0205 | 3.1060 | 1.1836 | 0.7953 |
|        |                  | variance    | 0.2059 | 5.5199 | 3.4616 | 1.3431 | 0.0885 | 5.4283 | 3.2218 | 1.4555 | 0.0885 | 5.1999 | 3.3084 |
|        |                  | $|\text{skewness}|$   | 0.0804 | 9.7994 | 1.6360 | 1.3264 | 0.0014 | 15.9908 | 1.9602 | 1.6250 | 0.0004 | 15.9908 | 1.8430 |
| F14    | $idfscore_{1} (\text{noun})$ | 0.0002 | 0.0702 | 0.0058 | 0.0072 | 0.0002 | 0.1146 | 0.0050 | 0.0080 | 0.0002 | 0.1146 | 0.0053 | 0.0077 |
|        |                  | variance    | 0.0000 | 0.0087 | 0.0002 | 0.0006 | 0.0000 | 0.0246 | 0.0002 | 0.0008 | 0.0000 | 0.0246 | 0.0002 |
|        |                  | $|\text{skewness}|$   | 1.1309 | 18.2918 | 5.7815 | 3.0954 | 0.5637 | 19.0358 | 6.2418 | 3.4033 | 0.5637 | 19.0358 | 6.0755 |
|        | $idfscore_{2} (\text{noun})$ | 0.0000 | 0.0309 | 0.0040 | 0.0044 | 0.0000 | 0.0422 | 0.0035 | 0.0045 | 0.0000 | 0.0422 | 0.0037 | 0.0045 |
|        |                  | variance    | 0.0000 | 0.0037 | 0.0002 | 0.0004 | 0.0000 | 0.0046 | 0.0002 | 0.0004 | 0.0000 | 0.0046 | 0.0002 |
Table A4. Descriptive statistics on the features of F2 from the test bed.

| Type     | Features                               | SocialTERMs                       | EventTERMs                       | SocialTERMS + EventTERMs                       |
|----------|----------------------------------------|------------------------------------|----------------------------------|-----------------------------------------------|
|          | F21 featuresentscore(noun, pos = verb) | Min 0.0000 Max 0.8750 Mean 0.0186 SD 0.0678 | Min 0.0000 Max 0.6250 Mean 0.0054 SD 0.0319 | Min 0.0000 Max 0.8750 Mean 0.0102 SD 0.0485 |
|          | featuresentscore (noun, pos = adverb)  | -                                  | -                                | -                                              |
|          | featuresentscore (noun, pos = adjective)| -                                 | -                                | -                                              |
|          | nounsentscore(noun)                    | 0.2754 0.9227 0.3576 0.0451 0.2037 0.6793 0.3535 0.0402 0.2037 0.8927 0.3550 0.0421 | -                                | -                                              |
|          | newsentscore(news)                     | 0.0048 0.7668 0.1391 0.0531 0.0228 0.6672 0.0370 0.0442 0.0228 0.7668 0.0378 0.0476 | -                                | -                                              |
|          | mean                                   | 0.0335 18.6163 2.7072 1.8694 0.0442 17.8336 2.5756 1.8807 0.0442 18.6163 2.6231 1.8777 | -                                | -                                              |
|          | variance                                | 0.2624 0.5257 0.3573 0.0325 0.2365 0.6223 0.3564 0.0381 0.2365 0.6223 0.3567 0.0361 | -                                | -                                              |
|          | l skewness                              | 0.0000 0.3164 0.0129 0.0215 0.0000 0.6035 0.0149 0.0346 0.0000 0.6035 0.0142 0.0305 | -                                | -                                              |
|          | F23 sentimentscore2(noun)               | -                                  | -                                | -                                              |
|          | topicsentscore(topic)                   | -                                  | -                                | -                                              |
|          | mean                                   | 0.0000 8.5184 1.5585 1.2550 0.0000 9.6439 1.5191 1.2559 0.0000 9.6439 1.5333 1.2557 | -                                | -                                              |
|          | variance                                | 0.0000 0.8750 0.0186 0.0678 0.0000 0.6250 0.0054 0.0319 0.0000 0.8750 0.0102 0.0485 | -                                | -                                              |
|          | l skewness                              | 0.0000 0.0217 0.0085 0.1647 0.0000 0.7500 0.0294 0.0936 0.0000 0.8125 0.0507 0.1273 | -                                | -                                              |

Table A5. Descriptive statistics on the features of F3 from the test bed.

| Type     | Features                               | SocialTERMs                       | EventTERMs                       | SocialTERMS + EventTERMs                       |
|----------|----------------------------------------|------------------------------------|----------------------------------|-----------------------------------------------|
|          | F31 degree(noun, CBN^{co-news})        | Min 0.0000 Max 0.3314 Mean 0.0115 SD 0.0233 | Min 0.0000 Max 0.3808 Mean 0.0094 SD 0.0261 | Min 0.0000 Max 0.3808 Mean 0.0102 SD 0.0252 |
|          | closeness(noun, CBN^{co-news})         | -                                  | -                                | -                                              |
|          | betweenness(noun, CBN^{co-news})       | -                                  | -                                | -                                              |
|          | F32 degree(noun, CBN^{co-topic})       | -                                  | -                                | -                                              |
|          | closeness(noun, CBN^{co-topic})        | -                                  | -                                | -                                              |
|          | betweenness(noun, CBN^{co-topic})      | -                                  | -                                | -                                              |
|          | F33 degree(noun, ITN^{co-news}(topic)) | -                                  | -                                | -                                              |
|          | closeness(noun, ITN^{co-news}(topic))  | -                                  | -                                | -                                              |
|          | betweenness(noun, ITN^{co-news}(topic))| -                                  | -                                | -                                              |
|          | F34 degree(noun, ICN^{co-topic}(community)) | -                  | -                                | -                                              |
|          | closeness(noun, ICN^{co-topic}(community)) | -                  | -                                | -                                              |
|          | betweenness(noun, ICN^{co-topic}(community)) | -                  | -                                | -                                              |
Table A6. Pairwise \( t \) tests on three performance measures for adding a feature set.

(a) Performance Measure = Accuracy

| Hypothesis          | DT       |
|---------------------|----------|
|                     | BL DT    | Bagging DT | Boosting DT | RS DT    |
| \( H_0: F_1 > F_1 \) | \( t \)  | \( p \)    | \( t \)  | \( p \)  | \( t \)  | \( p \)  | \( t \)  | \( p \)  |
| \( F_1 + F_1 > F_1 \) | 0.2235   | 0.8240    | 6.2247   | 0.0000   | 7.1741   | 0.0000   | 9.7341   | 0.0000   |
| \( F_2 + F_2 > F_2 \) | 30.8821  | 0.0000    | 37.4364  | 0.0000   | 21.4016  | 0.0000   | 28.7831  | 0.0000   |
| \( F_3 + F_3 > F_3 \) | 1.1166   | 0.2688    | 3.2303   | 0.0021   | 5.8678   | 0.0000   | 18.2384  | 0.0000   |

(b) Performance measure = \( F \)-measure

| Hypothesis          | NB       |
|---------------------|----------|
|                     | Bagging NB | Boosting NB | RS NB    |
| \( H_0: F_1 > F_1 \) | \( t \)  | \( p \)    | \( t \)  | \( p \)  | \( t \)  | \( p \)  | \( t \)  | \( p \)  |
| \( F_1 + F_1 > F_1 \) | 25.1039  | 0.0000    | 36.6398  | 0.0000   | 7.6015   | 0.0000   | 33.7182  | 0.0000   |
| \( F_2 + F_2 > F_2 \) | 19.9001  | 0.0000    | 42.3185  | 0.0000   | 31.3028  | 0.0000   | 47.0558  | 0.0000   |
| \( F_3 + F_3 > F_3 \) | 3.0525   | 0.0035    | -0.3495  | 0.7280   | 4.2665   | 0.0001   | -0.1539  | 0.8783   |

| Hypothesis          | RBFN     |
|---------------------|----------|
|                     | Bagging RBFN | Boosting RBFN | RS RBFN |
| \( H_0: F_1 > F_1 \) | \( t \)  | \( p \)    | \( t \)  | \( p \)  | \( t \)  | \( p \)  | \( t \)  | \( p \)  |
| \( F_1 + F_1 > F_1 \) | 47.8430  | 0.0000    | 26.1093  | 0.0000   | 24.5362  | 0.0000   | 39.7971  | 0.0000   |
| \( F_2 + F_2 > F_2 \) | 46.0717  | 0.0000    | 62.9056  | 0.0000   | 48.7427  | 0.0000   | 56.2702  | 0.0000   |
| \( F_3 + F_3 > F_3 \) | 1.6093   | 0.1131    | 12.8153  | 0.0000   | 7.9037   | 0.0000   | 11.2289  | 0.0000   |

| Hypothesis          | SVM      |
|---------------------|----------|
|                     | Bagging SVM | Boosting SVM | RS SVM   |
| \( H_0: F_1 > F_1 \) | \( t \)  | \( p \)    | \( t \)  | \( p \)  | \( t \)  | \( p \)  | \( t \)  | \( p \)  |
| \( F_1 + F_1 > F_1 \) | 1.7887   | 0.0791    | 4.1823   | 0.0001   | 4.7798   | 0.0000   | 2.0736   | 0.0427   |
| \( F_2 + F_2 > F_2 \) | 11.1502  | 0.0000    | 19.1245  | 0.0000   | 12.8451  | 0.0000   | 20.2212  | 0.0000   |
| \( F_3 + F_3 > F_3 \) | 1.4964   | 0.1401    | 4.3954   | 0.0001   | 0.8260   | 0.4123   | 1.9701   | 0.0537   |

| Hypothesis          | DBN      |
|---------------------|----------|
|                     | Bagging DBN | Boosting DBN | RS DBN   |
| \( H_0: F_1 > F_1 \) | \( t \)  | \( p \)    | \( t \)  | \( p \)  | \( t \)  | \( p \)  | \( t \)  | \( p \)  |
| \( F_1 + F_1 > F_1 \) | 1.7887   | 0.0791    | 4.1823   | 0.0001   | 4.7798   | 0.0000   | 2.0736   | 0.0427   |
| \( F_2 + F_2 > F_2 \) | 11.1502  | 0.0000    | 19.1245  | 0.0000   | 12.8451  | 0.0000   | 20.2212  | 0.0000   |
| \( F_3 + F_3 > F_3 \) | 1.4964   | 0.1401    | 4.3954   | 0.0001   | 0.8260   | 0.4123   | 1.9701   | 0.0537   |
| Table A6. Cont. |
|----------------|
| (b) Performance measure = F-measure |
| RBFN | BL RBFN | Bagging RBFN | Boosting RBFN | RS RBFN |
| t | p | t | p | t | p | t | p |
| F1 + F1 (>) F1 (>) | -8.9487 | 0.0000 | 2.7254 | 0.0141 | 21.9769 | 0.0000 | 2.6784 | 0.0162 |
| F2 + F2 (>) F2 (>) | 18.8657 | 0.0000 | 18.5574 | 0.0000 | 39.4377 | 0.0000 | 16.4852 | 0.0000 |
| F3 + F3 (>) F3 (>) | 1.1519 | 0.2644 | 0.3593 | 0.7237 | 22.9424 | 0.0000 | 0.5061 | 0.6189 |
| SVM | BL SVM | Bagging SVM | Boosting SVM | RS SVM |
| t | p | t | p | t | p | t | p |
| F1 + F1 (>) F1 (>) | 51.3223 | 0.0000 | 20.3092 | 0.0000 | 12.6380 | 0.0000 | 28.9782 | 0.0000 |
| F2 + F2 (>) F2 (>) | 15.6709 | 0.0000 | 26.6824 | 0.0000 | 27.7334 | 0.0000 | 35.0939 | 0.0000 |
| F3 + F3 (>) F3 (>) | 1.5515 | 0.1400 | 2.5695 | 0.0197 | 0.5061 | 0.6189 | 0.9683 | 0.3458 |
| DBN | BL DBN | Bagging DBN | Boosting DBN | RS DBN |
| t | p | t | p | t | p | t | p |
| F1 + F1 (>) F1 (>) | 0.2507 | 0.8049 | 1.7329 | 0.1056 | 3.0737 | 0.0066 | 1.1969 | 0.2469 |
| F2 + F2 (>) F2 (>) | 1.5755 | 0.1357 | 2.0121 | 0.0600 | 4.2951 | 0.0004 | 6.9531 | 0.0000 |
| F3 + F3 (>) F3 (>) | 0.0878 | 0.9310 | 2.5695 | 0.0197 | 0.8361 | 0.4143 | 0.9683 | 0.3458 |
| Hypothesis | DT | BL DT | Bagging DT | Boosting DT | RS DT |
| t | p | t | p | t | p | t | p |
| F1 + F1 (>) F1 (>) | -5.2998 | 0.0000 | 11.7273 | 0.0000 | 8.3059 | 0.0000 | 14.2223 | 0.0000 |
| F2 + F2 (>) F2 (>) | 17.5129 | 0.0000 | 35.0495 | 0.0000 | 16.1747 | 0.0000 | 31.3389 | 0.0000 |
| F3 + F3 (>) F3 (>) | 4.3007 | 0.0001 | 2.9594 | 0.0045 | 24.0347 | 0.0000 | 2.2664 | 0.0272 |
| NB | BL NB | Bagging NB | Boosting NB | RS NB |
| t | p | t | p | t | p | t | p |
| F1 + F1 (>) F1 (>) | 57.1072 | 0.0000 | 11.3264 | 0.0000 | 18.7506 | 0.0000 | 5.5378 | 0.0000 |
| F2 + F2 (>) F2 (>) | 68.7347 | 0.0000 | 88.4192 | 0.0000 | 64.7801 | 0.0000 | 93.4364 | 0.0000 |
| F3 + F3 (>) F3 (>) | 1.9929 | 0.0510 | -2.7524 | 0.0081 | 9.3195 | 0.0000 | -8.1966 | 0.0000 |
| RBFN | BL RBFN | Bagging RBFN | Boosting RBFN | RS RBFN |
| t | p | t | p | t | p | t | p |
| F1 + F1 (>) F1 (>) | -5.2998 | 0.0000 | 10.7145 | 0.0000 | 43.9045 | 0.0000 | 9.4867 | 0.0000 |
| F2 + F2 (>) F2 (>) | 17.5129 | 0.0000 | 42.0800 | 0.0000 | 72.7003 | 0.0000 | 36.3416 | 0.0000 |
| F3 + F3 (>) F3 (>) | 4.3007 | 0.0001 | 2.9594 | 0.0045 | 24.0347 | 0.0000 | 2.2664 | 0.0272 |
| SVM | BL SVM | Bagging SVM | Boosting SVM | RS SVM |
| t | p | t | p | t | p | t | p |
| F1 + F1 (>) F1 (>) | 50.6939 | 0.0000 | 37.2924 | 0.0000 | 32.2763 | 0.0000 | 41.9019 | 0.0000 |
| F2 + F2 (>) F2 (>) | 40.2739 | 0.0000 | 78.4504 | 0.0000 | 52.7703 | 0.0000 | 66.4601 | 0.0000 |
| F3 + F3 (>) F3 (>) | 1.8869 | 0.0644 | 21.8909 | 0.0000 | 11.0255 | 0.0000 | 11.4046 | 0.0000 |
| DBN | BL DBN | Bagging DBN | Boosting DBN | RS DBN |
| t | p | t | p | t | p | t | p |
| F1 + F1 (>) F1 (>) | 1.7315 | 0.0889 | 4.5510 | 0.0000 | 5.0435 | 0.0000 | 1.6871 | 0.0970 |
| F2 + F2 (>) F2 (>) | 11.7540 | 0.0000 | 21.8630 | 0.0000 | 12.5776 | 0.0000 | 20.4907 | 0.0000 |
| F3 + F3 (>) F3 (>) | 1.4243 | 0.1598 | 4.7451 | 0.0000 | 0.9398 | 0.3513 | 1.7289 | 0.0892 |

Notes: F1 (>) = F2 + F3, F2 (>) = F1 + F3, and F3 (>) = F1 + F2. The results are t and p values of the t tests for feature set comparisons, and the results with a significance level higher than 5% are italicized.
Table A7. Pairwise t tests on three performance measures for different base learners.

(a) Performance Measure = Accuracy

| Hypothesis          | F1       | F1 + F2   | F1 + F2 + F3 |
|---------------------|----------|-----------|--------------|
|                     | t        | p         | t            | p           | t            | p           |
| BL NB > BL DT       | 12.0713  | 0.0000    | -69.5140     | 0.0000      | -72.8854     | 0.0000      |
| BL RBFN > BL DT     | 0.0000   | 1.0000    | 0.0000       | 1.0000      | 0.0000       | 1.0000      |
| BL SVM > BL DT      | 3.4122   | 0.0015    | -37.1666     | 0.0000      | -36.7021     | 0.0000      |
| BL DBM > BL DT      | -16.3951 | 0.0000    | -19.0950     | 0.0000      | -20.0959     | 0.0000      |
| BL RBFN > BL NB     | -12.0713 | 0.0000    | 69.5140      | 0.0000      | 72.8854      | 0.0000      |
| BL SVM > BL NB      | -14.1094 | 0.0000    | 45.4236      | 0.0000      | 41.6482      | 0.0000      |
| BL DBN > BL RBFN    | -21.8254 | 0.0000    | -1.0832      | 0.2874      | 0.5031       | 0.6186      |
| BL SVM > BL DBM     | -15.0229 | 0.0000    | -35.9530     | 0.0000      | -34.3491     | 0.0000      |
| BL DBM > BL SVM     | -20.0959 | 0.0000    | -9.2654      | 0.0000      | -8.9652      | 0.0000      |

(b) Performance measure = F-measure

| Hypothesis          | F1       | F1 + F2   | F1 + F2 + F3 |
|---------------------|----------|-----------|--------------|
|                     | t        | p         | t            | p           | t            | p           |
| BL NB > BL DT       | 17.3022  | 0.0000    | -104.7576    | 0.0000      | -84.9164     | 0.0000      |
| BL RBFN > BL DT     | 0.0000   | 1.0000    | 0.0000       | 1.0000      | 0.0000       | 1.0000      |
| BL SVM > BL DT      | 5.1433   | 0.0000    | -49.4801     | 0.0000      | -37.6746     | 0.0000      |
| BL DBM > BL DT      | -18.9402 | 0.0000    | -19.1557     | 0.0000      | -20.0682     | 0.0000      |
| BL RBFN > BL NB     | -17.3022 | 0.0000    | 104.7576     | 0.0000      | 84.9164      | 0.0000      |
| BL SVM > BL NB      | -25.8195 | 0.0000    | 85.0502      | 0.0000      | 68.8742      | 0.0000      |
| BL DBN > BL RBFN    | -23.0363 | 0.0000    | -4.8859      | 0.0000      | -4.1743      | 0.0000      |
| BL SVM > BL DBM     | 5.1433   | 0.0000    | -49.4801     | 0.0000      | -37.6746     | 0.0000      |
| BL DBN > BL SVM     | -18.4671 | 0.0000    | -9.2654      | 0.0000      | -8.9652      | 0.0000      |

(c) Performance measure = AUC

| Hypothesis          | F1       | F1 + F2   | F1 + F2 + F3 |
|---------------------|----------|-----------|--------------|
|                     | t        | p         | t            | p           | t            | p           |
| BL NB > BL DT       | 21.2132  | 0.0000    | 6.1123       | 0.0000      | 1.9996       | 0.0535      |
| BL RBFN > BL DT     | 0.0000   | 1.0000    | 0.0000       | 1.0000      | 0.0000       | 1.0000      |
| BL SVM > BL DT      | -15.0229 | 0.0000    | -35.9530     | 0.0000      | -34.3491     | 0.0000      |
| BL DBM > BL DT      | -25.1656 | 0.0000    | -24.8453     | 0.0000      | -27.2766     | 0.0000      |
| BL RBFN > BL NB     | -21.2132 | 0.0000    | -6.1123      | 0.0000      | -1.9996      | 0.0535      |
| BL SVM > BL NB      | -80.0403 | 0.0000    | -86.1265     | 0.0000      | -99.1118     | 0.0000      |
| BL DBN > BL NB      | -38.3070 | 0.0000    | -30.1059     | 0.0000      | -33.1400     | 0.0000      |
| BL SVM > BL RBFN    | -15.0229 | 0.0000    | -35.9530     | 0.0000      | -34.3491     | 0.0000      |
| BL DBN > BL RBFN    | -25.1656 | 0.0000    | -24.8453     | 0.0000      | -27.2766     | 0.0000      |
| BL DBN > BL SVM     | -19.7493 | 0.0000    | -9.7627      | 0.0000      | -9.7155      | 0.0000      |

Notes: The results are t and p values of the t tests for classification technique comparisons, and the results with a significance level higher than 5% are italicized.
Table A8. Pairwise t tests on three performance measures for base leaner vs. ensemble learning methods.

(a) Performance Measure = Accuracy

| Ensemble methods | Hypothesis         | F1     | F1 + F2 | F1 + F2 + F3 |
|------------------|--------------------|--------|---------|--------------|
|                  |                    | t      | p       | t           | p           |
| Bagging          | Bagging DT > BL DT | 88.4498 | 0.0000  | 106.0151    | 0.0000      |
|                  | Bagging NB > BL NB | −17.7052 | 0.0000  | 27.1656     | 0.0000      |
|                  | Bagging RBFN > BL RBFN | 67.7853 | 0.0000  | 81.6247     | 0.0000      |
|                  | Bagging SVM > BL SVM | 2.7899  | 0.0074  | 15.6135     | 0.0000      |
|                  | Bagging DBN > BL DBN | 2.8758  | 0.0056  | 3.7963      | 0.0004      |
| Boosting         | Boosting DT > BL DT | 14.4893 | 0.0000  | 71.1436     | 0.0000      |
|                  | Boosting NB > BL NB | −6.5743  | 0.0000  | 30.5947     | 0.0000      |
|                  | Boosting RBFN > BL RBFN | 5.1906  | 0.0000  | −3.8927     | 0.0003      |
|                  | Boosting SVM > BL SVM | 2.7899  | 0.0074  | 15.6135     | 0.0000      |
|                  | Boosting DBN > BL DBN | 2.8758  | 0.0056  | 3.7963      | 0.0004      |
| RS               | RS DT > BL DT      | 10.7463 | 0.0000  | 20.2510     | 0.0000      |
|                  | RS NB > BL NB      | −20.3202 | 0.0000  | 32.3685     | 0.0000      |
|                  | RS RBFN > BL RBFN | 87.0035 | 0.0000  | 89.3507     | 0.0000      |
|                  | RS SVM > BL SVM    | 2.1645  | 0.0347  | 22.0020     | 0.0000      |
|                  | RS DBN > BL DBN    | −10.7311 | 0.0000  | −0.3488     | 0.7285      |

(b) Performance measure = F-measure

| Ensemble methods | Hypothesis         | F1     | F1 + F2 | F1 + F2 + F3 |
|------------------|--------------------|--------|---------|--------------|
|                  |                    | t      | p       | t           | p           |
| Bagging          | Bagging DT > BL DT | 75.9369 | 0.0000  | 132.9017    | 0.0000      |
|                  | Bagging NB > BL NB | −21.5170 | 0.0000  | 32.3685     | 0.0000      |
|                  | Bagging RBFN > BL RBFN | 66.4283 | 0.0000  | 81.7851     | 0.0000      |
|                  | Bagging SVM > BL SVM | 3.8242  | 0.0003  | 4.4323      | 0.0001      |
|                  | Bagging DBN > BL DBN | 2.9466  | 0.0048  | −5.3816     | 0.0000      |
| Boosting         | Boosting DT > BL DT | 18.3200 | 0.0000  | 73.1555     | 0.0000      |
|                  | Boosting NB > BL NB | −8.7085  | 0.0000  | 68.8445     | 0.0000      |
|                  | Boosting RBFN > BL RBFN | 2.9466  | 0.0048  | −5.3816     | 0.0000      |
|                  | Boosting SVM > BL SVM | 9.5777  | 0.0005  | 3.7196      | 0.0005      |
|                  | Boosting DBN > BL DBN | 6.7542  | 0.0000  | 23.6999     | 0.0000      |
| RS               | RS DT > BL DT      | 6.7542  | 0.0000  | 23.6999     | 0.0000      |
|                  | RS NB > BL NB      | −24.8591 | 0.0000  | 98.7848     | 0.0000      |
|                  | RS RBFN > BL RBFN | 11.5363 | 0.0000  | 22.3686     | 0.0000      |
|                  | RS SVM > BL SVM    | 1.8305  | 0.0725  | 36.7534     | 0.0000      |
|                  | RS DBN > BL DBN    | −13.2600 | 0.0000  | −0.4610     | 0.6466      |

(c) Performance measure = AUC

| Ensemble methods | Hypothesis         | F1     | F1 + F2 | F1 + F2 + F3 |
|------------------|--------------------|--------|---------|--------------|
|                  |                    | t      | p       | t           | p           |
| Bagging          | Bagging DT > BL DT | 120.5013 | 0.0000  | 112.7957     | 0.0000      |
|                  | Bagging NB > BL NB | −29.5658 | 0.0000  | −31.5824     | 0.0000      |
|                  | Bagging RBFN > BL RBFN | 95.6464 | 0.0000  | 90.0383     | 0.0000      |
|                  | Bagging SVM > BL SVM | 28.8670 | 0.0000  | 66.8031     | 0.0000      |
|                  | Bagging DBN > BL DBN | 4.1559  | 0.0001  | 4.7537      | 0.0000      |
| Boosting         | Boosting DT > BL DT | 25.4924 | 0.0000  | 83.0114     | 0.0000      |
|                  | Boosting NB > BL NB | −15.7358 | 0.0000  | −8.5868     | 0.0000      |
|                  | Boosting RBFN > BL RBFN | 11.5363 | 0.0000  | 22.3686     | 0.0000      |
|                  | Boosting SVM > BL SVM | 29.5838 | 0.0000  | 77.8730     | 0.0000      |
|                  | Boosting DBN > BL DBN | 7.9322  | 0.0000  | 1.8372      | 0.0713      |
| RS               | RS DT > BL DT      | 20.8164 | 0.0000  | 34.0770     | 0.0000      |
|                  | RS NB > BL NB      | −28.0384 | 0.0000  | −33.2338     | 0.0000      |
|                  | RS RBFN > BL RBFN | 100.5234 | 0.0000  | 97.6323     | 0.0000      |
|                  | RS SVM > BL SVM    | 2.0360  | 0.0467  | 20.9832     | 0.0000      |
|                  | RS DBN > BL DBN    | −12.5805 | 0.0000  | −0.3533     | 0.7251      |

Notes: The results are t and p values of the t tests for classification technique comparisons, and the results with a significance level higher than 5% are italicized.
Table A9. Pairwise *t* tests on three performance measures for different ensemble methods, i.e., ensemble method vs. ensemble method.

(a) Performance Measure = Accuracy

| Base learners | Hypothesis                         | F1         | F1 + F2     | F1 + F2 + F3    |
|---------------|-----------------------------------|------------|-------------|-----------------|
|               |                                   | *t*        | *p*         | *t*            | *p*            |
| DT            | Boosting DT > Bagging DT          | −34.7363   | 0.0000      | 5.7999         | 0.0000         |
|               | RS DT > Bagging DT                | −72.3451   | 0.0000      | −43.9322       | 0.0000         |
|               | RS DT > Boosting DT               | −8.3594    | 0.0000      | −39.6975       | 0.0000         |
| NB            | Boosting NB > Bagging NB          | 8.2097     | 0.0000      | 12.8248        | 0.0000         |
|               | RS NB > Bagging NB                | −3.6472    | 0.0066      | 1.7535         | 0.0850         |
|               | RS NB > Boosting NB               | −10.6185   | 0.0000      | −12.1853       | 0.0000         |
| RBFN          | Boosting RBFN > Bagging RBFN      | −64.3608   | 0.0000      | −72.3535       | 0.0000         |
|               | RS RBFN > Bagging RBFN            | 8.1792     | 0.0000      | 9.4283         | 0.0000         |
|               | RS RBFN > Boosting RBFN           | 83.6084    | 0.0000      | 79.5634        | 0.0000         |
| SVM           | Boosting SVM > Bagging SVM        | −0.9735    | 0.3345      | 2.4655         | 0.0167         |
|               | RS SVM > Bagging SVM              | −0.7949    | 0.4300      | 1.4032         | 0.1666         |
|               | RS SVM > Boosting SVM             | 0.1963     | 0.8450      | −1.4398        | 0.1597         |
| DBN           | Boosting DBN > Bagging DBN        | 7.2925     | 0.0000      | 0.3045         | 0.7619         |
|               | RS DBN > Bagging DBN              | −13.8808   | 0.0000      | −4.0145        | 0.0002         |
|               | RS DBN > Boosting DBN             | −30.0740   | 0.0000      | −4.0145        | 0.0002         |

(b) Performance measure = F-measure

| Base learners | Hypothesis                         | F1         | F1 + F2     | F1 + F2 + F3    |
|---------------|-----------------------------------|------------|-------------|-----------------|
|               |                                   | *t*        | *p*         | *t*            | *p*            |
| DT            | Boosting DT > Bagging DT          | −32.4298   | 0.0000      | 4.1017         | 0.0002         |
|               | RS DT > Bagging DT                | −57.8336   | 0.0000      | −47.7845       | 0.0000         |
|               | RS DT > Boosting DT               | −12.0562   | 0.0000      | −39.2425       | 0.0000         |
| NB            | Boosting NB > Bagging NB          | 8.2602     | 0.0000      | 14.2373        | 0.0000         |
|               | RS NB > Bagging NB                | −1.9306    | 0.0856      | 0.5775         | 0.5665         |
|               | RS NB > Boosting NB               | −9.9778    | 0.0000      | −15.6158       | 0.0000         |
| RBFN          | Boosting RBFN > Bagging RBFN      | −76.5938   | 0.0000      | −78.7143       | 0.0000         |
|               | RS RBFN > Bagging RBFN            | 8.0850     | 0.0000      | 7.5814         | 0.0000         |
|               | RS RBFN > Boosting RBFN           | 76.6385    | 0.0000      | 80.7990        | 0.0000         |
| SVM           | Boosting SVM > Bagging SVM        | −1.3708    | 0.1772      | 1.8589         | 0.0682         |
|               | RS SVM > Bagging SVM              | −1.3800    | 0.1747      | −1.3300        | 0.1893         |
|               | RS SVM > Boosting SVM             | 0.0280     | 0.9778      | −3.6290        | 0.0006         |
| DBN           | Boosting DBN > Bagging DBN        | 5.6423     | 0.0000      | −0.2732        | 0.7588         |
|               | RS DBN > Bagging DBN              | −19.3905   | 0.0000      | −4.8681        | 0.0000         |
|               | RS DBN > Boosting DBN             | −31.3674   | 0.0000      | −4.1374        | 0.0001         |

(c) Performance measure = AUC

| Base learners | Hypothesis                         | F1         | F1 + F2     | F1 + F2 + F3    |
|---------------|-----------------------------------|------------|-------------|-----------------|
|               |                                   | *t*        | *p*         | *t*            | *p*            |
| DT            | Boosting DT > Bagging DT          | −34.6090   | 0.0000      | 8.0364         | 0.0000         |
|               | RS DT > Bagging DT                | −78.7128   | 0.0000      | −52.1042       | 0.0000         |
|               | RS DT > Boosting DT               | −12.0881   | 0.0000      | −45.2120       | 0.0000         |
| NB            | Boosting NB > Bagging NB          | 0.8335     | 0.4099      | 23.9324        | 0.0000         |
|               | RS NB > Bagging NB                | 0.6667     | 0.5076      | −4.8801        | 0.0000         |
|               | RS NB > Boosting NB               | −0.4738    | 0.6383      | −26.0628       | 0.0000         |
| RBFN          | Boosting RBFN > Bagging RBFN      | −98.5728   | 0.0000      | −104.3890      | 0.0000         |
|               | RS RBFN > Bagging RBFN            | 10.1187    | 0.0000      | 8.9604         | 0.0000         |
|               | RS RBFN > Boosting RBFN           | 103.1878   | 0.0000      | 117.4996       | 0.0000         |
| SVM           | Boosting SVM > Bagging SVM        | 5.2714     | 0.0000      | 9.4300         | 0.0000         |
|               | RS SVM > Bagging SVM              | −29.6967   | 0.0000      | −43.9027       | 0.0000         |
|               | RS SVM > Boosting SVM             | −29.9190   | 0.0000      | −54.0403       | 0.0000         |
| DBN           | Boosting DBN > Bagging DBN        | 3.7328     | 0.0004      | −2.7483        | 0.0084         |
|               | RS DBN > Bagging DBN              | −18.9113   | 0.0000      | −4.9850        | 0.0000         |
|               | RS DBN > Boosting DBN             | −25.4271   | 0.0000      | −2.1488        | 0.0359         |

Notes: The results are *t* and *p* values of the *t* tests for classification technique comparisons, and the results with a significance level higher than 5% are italicized.
References

1. Myung, W.; Lee, G.-H.; Won, H.-H.; Fava, M.; Mischoulon, D.; Nyer, M.; Kim, D.K.; Heo, J.-Y.; Jeon, H.J. Paraquat Prohibition and Change in the Suicide Rate and Methods in South Korea. *PLoS ONE* 2015, *10*, e0128980. [CrossRef] [PubMed]

2. Ittipanuvat, V.; Fujita, K.; Sakata, I.; Kajikawa, Y. Finding linkage between technology and social issue: A Literature Based Discovery approach. *J. Eng. Technol. Manag.* 2014, *32* (Suppl. C), 160–184. [CrossRef]

3. Phillips, W.; Lee, H.; Ghobadian, A.; O’Regan, N.; James, P. Social Innovation and Social Entrepreneurship: A Systematic Review. *Group Organ. Manag.* 2015, *40*, 428–461. [CrossRef]

4. Chang, R.M.; Kauffman, R.J.; Kwon, Y. Understanding the paradigm shift to computational social science in the presence of big data. *Decis. Support Syst.* 2014, *63* (Suppl. C), 67–80. [CrossRef]

5. Chi, Y.L. A consumer-centric design approach to develop comprehensive knowledge-based systems for keyword discovery. *Exp. Syst. Appl.* 2009, *36*, 2481–2493. [CrossRef]

6. Jiang, S.; Chen, H.; Nunamaker, J.F.; Zimbra, D. Analyzing firm-specific social media and market: A stakeholder-based event analysis framework. *Decis. Support Syst.* 2014, *67*, 30–39. [CrossRef]

7. Zhang, Y.L.; Dang, Y.; Chen, H.C. Gender Classification for Web Forums. *IEEE Trans. Syst. Man Cybern. A* 2011, *41*, 668–677. [CrossRef]

8. Suh, J.H. Forecasting the daily outbreak of topic-level political risk from social media using hidden Markov model-based techniques. *Technol. Forecast. Soc. Change* 2015, *94*, 115–132. [CrossRef]

9. Suh, J.H.; Park, C.H.; Jeon, S.H. Applying text and data mining techniques to forecasting the trend of petitions filed to e-People. *Exp. Syst. Appl.* 2010, *37*, 7255–7268. [CrossRef]

10. Einav, L.; Levin, J. Economics in the age of big data. *Science* 2014, *346*, 1243089. [CrossRef]

11. Debortoli, S.; Müller, O.; Junglas, I.A.; vom Brocke, J. Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial. *Commun. Assoc. Inf. Syst.* 2016, *39*, 110–135. [CrossRef]

12. Conde, A.; Larrañaga, M.; Arruarte, A.; Elorriaga, J.A.; Roth, D. litewi: A combined term extraction and entity linking method for eliciting educational ontologies from textbooks. *J. Assoc. Inf. Sci. Technol.* 2016, *67*, 380–399. [CrossRef]

13. Tseng, Y.-H.; Lin, C.-J.; Lin, Y.-I. Text mining techniques for patent analysis. *Inf. Proc. Manag.* 2007, *43*, 1216–1247. [CrossRef]

14. Chou, C.-H.; Sinha, A.P.; Zhao, H. Commercial Internet filters: Perils and opportunities. *Decis. Support Syst.* 2010, *48*, 521–530. [CrossRef]

15. Dang, Y.; Zhang, Y.L.; Chen, H.C.; Hu, P.J.H.; Brown, S.A.; Larson, C. Arizona Literature Mapper: An Integrated Approach to Monitor and Analyze Global Bioterrorism Research Literature. *J. Am. Soc. Inf. Sci. Technol.* 2009, *60*, 1466–1485. [CrossRef]

16. He, Q.; Chang, K.; Lim, E.P.; Banerjee, A. Keep it simple with time: A reexamination of probabilistic topic detection models. *IEEE Trans. Pattern Anal. Mach. Intell.* 2010, *32*, 1795–1808. [PubMed]

17. Hu, Y.-H.; Chen, Y.-L.; Chou, H.-L. Opinion mining from online hotel reviews—A text summarization approach. *Inf. Process. Manag.* 2017, *53*, 436–449. [CrossRef]

18. Liu, Y.-H.; Chen, Y.-L.; Ho, W.-L. Predicting associated statutes for legal problems. *Inf. Process. Manag.* 2015, *51*, 194–211. [CrossRef]

19. Suh, J.H. Comparing writing style feature-based classification methods for estimating user reputations in social media. *Springerplus* 2016, *5*, 261. [CrossRef]

20. Nieminen, P.; Polonen, I.; Sipola, T. Research literature clustering using diffusion maps. *J. Informetr.* 2013, *7*, 874–886. [CrossRef]

21. Chen, K.-Y.; Luesukprasert, L.; Chou, S.-C.T. Hot Topic Extraction Based on Timeline Analysis and Multidimensional Sentence Modeling. *IEEE Trans. Knowl. Data Eng.* 2007, *19*, 1016–1025. [CrossRef]

22. Xu, H.; Zhang, F.; Wang, W. Implicit feature identification in Chinese reviews using explicit topic mining model. *Knowl.-Based Syst.* 2015, *76*, 166–173. [CrossRef]

23. Zheng, X.L.; Lin, Z.; Wang, X.W.; Lin, K.J.; Song, M.N. Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification. *Knowl.-Based Syst.* 2014, *61*, 29–47. [CrossRef]

24. Rose, S.; Engel, D.; Cramer, N.; Cowley, W. Automatic keyword extraction from individual documents. *Text Min.* 2010, *1–20*. 

Sustainability 2019, 11, 196
25. Abilhoa, W.D.; de Castro, L.N. A keyword extraction method from twitter messages represented as graphs. 
   Appl. Math. Comput. 2014, 240, 308–325. [CrossRef]
26. Noh, H.; Jo, Y.; Lee, S. Keyword selection and processing strategy for applying text mining to patent analysis. 
   Exp. Syst. Appl. 2015, 42, 4348–4360. [CrossRef]
27. Piriyani, R.; Madhavi, D.; Singh, V.K. Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. 
   Inf. Process. Manag. 2017, 53, 122–150. [CrossRef]
28. Yang, S.; Han, R.; Wolfram, D.; Zhao, Y. Visualizing the intellectual structure of information science (2006–2015): Introducing author keyword coupling analysis. 
   J. Informetr. 2016, 10, 132–150. [CrossRef]
29. Liu, Z.; Jansen, B.J. Questioner or question: Predicting the response rate in social question and answering on Sina Weibo. 
   Inf. Process. Manag. 2018, 54, 159–174. [CrossRef]
30. Peetz, M.-H.; de Rijke, M.; Kaptein, R. Estimating Reputation Polarity on Microblog Posts. 
   Inf. Process. Manag. 2016, 52, 193–216. [CrossRef]
31. Almeida, T.A.; Silva, T.P.; Santos, I.; Gómez Hidalgo, J.M. Text normalization and semantic indexing to enhance Instant Messaging and SMS spam filtering. 
   Knowl.-Based Syst. 2016, 108 (Suppl. C), 25–32. [CrossRef]
32. Rao, Y.; Li, Q.; Wu, Q.; Xie, H.; Wang, F.L.; Wang, T. A multi-relational term scheme for first story detection. 
   Neurocomputing 2017, 254 (Suppl. C), 42–52. [CrossRef]
33. Lin, D.; Li, L.; Cao, D.; Lv, Y.; Ke, X. Multi-modality weakly labeled sentiment learning based on Explicit Emotion Signal for Chinese microblog. 
   Neurocomputing 2018, 272 (Suppl. C), 258–269. [CrossRef]
34. Alruily, M.; Ayesh, A.; Zedan, H. Crime profiling for the Arabic language using computational linguistic techniques. 
   Inf. Process. Manag. 2014, 50, 315–341. [CrossRef]
35. Lo, S.L.; Chiong, R.; Cornforth, D. An unsupervised multilingual approach for online social media topic identification. 
   Exp. Syst. Appl. 2017, 81 (Suppl. C), 282–298. [CrossRef]
36. Pournarakis, D.E.; Sotiropoulos, D.N.; Giaglis, G.M. A computational model for mining consumer perceptions in social media. 
   Decis. Support Syst. 2017, 93 (Suppl. C), 98–110. [CrossRef]
37. Zhang, Y.; Porter, A.L.; Hu, Z.; Guo, Y.; Newman, N.C. “Term clumping” for technical intelligence: A case study on dye-sensitized solar cells. 
   Technol. Forecast. Soc. Change 2014, 85, 26–39. [CrossRef]
38. Li, Q.; Jin, Z.; Wang, C.; Zeng, D.D. Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems. 
   Knowl.-Based Syst. 2016, 107 (Suppl. C), 289–300. [CrossRef]
39. Weichselbraun, A.; Gindl, S.; Scharl, A. Enriching semantic knowledge bases for opinion mining in big data applications. 
   Knowl.-Based Syst. 2014, 69, 78–85. [CrossRef]
40. Li, Q.; Liu, Y. Exploring the diversity of retweeting behavior patterns in Chinese microblogging platform. 
   Inf. Process. Manag. 2017, 53, 945–962. [CrossRef]
41. Jung, S.; Segev, A. Analyzing future communities in growing citation networks. 
   Knowl.-Based Syst. 2014, 69, 34–44. [CrossRef]
42. Lee, Y.; Kim, S.Y.; Song, I.; Park, Y.; Shin, J. Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis. 
   Scientometrics 2014, 100, 227–244. [CrossRef]
43. Dang, Y.; Zhang, Y.; Chen, H. A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews. 
   IEEE Intell. Syst. 2010, 25, 46–53. [CrossRef]
44. Chen, C.C.; Chen, Y.T.; Chen, M.C. An aging theory for event life-cycle modeling. 
   IEEE Trans. Syst. Man Cybern. A 2009, 39, 850–863. [CrossRef]
50. Schumaker, R.P.; Chen, H. Textual analysis of stock market prediction using breaking financial news. ACM Trans. Inf. Syst. 2009, 27, 1–19. [CrossRef]
51. Huang, H.-H.; Kuo, Y.-H. Cross-Lingual Document Representation and Semantic Similarity Measure: A Fuzzy Set and Rough Set Based Approach. IEEE Trans. Fuzzy Syst. 2010, 18, 1098–1111. [CrossRef]
52. Spina, D.; Gonzalo, J.; Amigó, E. Discovering filter keywords for company name disambiguation in twitter. Exp. Syst. Appl. 2013, 40, 4986–5003. [CrossRef]
53. Sheth, A.; Thomas, C.; Mehra, P. Continuous Semantics to Analyze Real-Time Data. IEEE Int. Comput. 2010, 14, 84–89. [CrossRef]
54. Jatowt, A.; Au Yeung, C.M.; Tanaka, K. Generic method for detecting focus time of documents. Decis. Support Syst. 2013, 55, 851–860. [CrossRef]
55. Zhang, Y.L.; Dang, Y.; Chen, H.C. Research note: Examining gender emotional differences in Web forum communication. Decis. Support Syst. 2015, 51, 851–868. [CrossRef]
56. Ji, X.; Chun, S.A.; Wei, Z.; Geller, J. Twitter sentiment classification for measuring public health concerns. Soc. Netw. Anal. Min. 2015, 5, 13. [CrossRef]
57. Borgatti, S.P.; Mehra, A.; Brass, D.J.; Labianca, G. Network Analysis in the Social Sciences. Science 2009, 323, 892–895. [CrossRef]
58. Yan, E.J.; Ding, Y. Applying Centrality Measures to Impact Analysis: A Coauthorship Network Analysis. J. Am. Soc. Inf. Sci. Technol. 2009, 60, 2107–2118. [CrossRef]
59. Suh, J.H. Exploring the effect of structural patent indicators in forward patent citation networks on patent price from firm market value. Technol. Anal. Strateg. Manag. 2015, 27, 485–502. [CrossRef]
60. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.U. Complex networks: Structure and dynamics. Phys. Rep. 2006, 424, 175–308. [CrossRef]
61. Akoglu, L.; Tong, H.; Koutra, D. Graph based anomaly detection and description: A survey. Data Min. Knowl. Discov. 2014, 29, 626–688. [CrossRef]
62. Steinhaeuser, K.; Chawla, N.V. Identifying and evaluating community structure in complex networks. Pattern Recognit. Lett. 2010, 31, 413–421. [CrossRef]
63. Zhao, Z.; Feng, S.; Wang, Q.; Huang, J.Z.; Williams, G.J.; Fan, J. Topic oriented community detection through social objects and link analysis in social networks. Knowl.-Based Syst. 2012, 26, 164–173. [CrossRef]
64. Expert, P.; Evans, T.S.; Blondel, V.D.; Lambiotte, R. Uncovering space-independent communities in spatial networks. Proc. Natl. Acad. Sci. USA 2011, 108, 7663–7668. [CrossRef] [PubMed]
65. Vincent, D.B.; Jean-Loup, G.; Renaud, L.; Etienne, L. Fast unfolding of communities in large networks. J. Statist. Mech. 2008, 2008, P10008.
66. Nettleton, D.F. Data mining of social networks represented as graphs. Comput. Sci. Rev. 2013, 7, 1–34. [CrossRef]
67. Ku, Y.; Chiu, C.; Zhang, Y.; Chen, H.; Su, H. Text mining self-disclosing health information for public health service. J. Assoc. Inf. Sci. Technol. 2014, 65, 928–947. [CrossRef]
68. Abbasi, A.; Chen, H.; Thoms, S.; Fu, T. Affect analysis of web forums and blogs using correlation ensembles. IEEE Trans. Knowl. Data Eng. 2008, 20, 1168–1180. [CrossRef]
69. Oztekin, A.; Kong, Z.Y.J.; Delen, D. Development of a structural equation modeling-based decision tree methodology for the analysis of lung transplantations. Decis. Support Syst. 2011, 51, 155–166. [CrossRef]
70. Yang, Y.M.; Slattery, S.; Ghani, R. A study of approaches to hypertext categorization. J. Intell. Inf. Syst. 2002, 18, 219–241. [CrossRef]
71. Roy, B.V.; Yan, X. Manipulation Robustness of Collaborative Filtering. Manag. Sci. 2010, 56, 1911–1929.
72. Witten, I.H.; Frank, E.; Hall, M.A. Data Mining: Practical Machine Learning Tools and Techniques; Practical Machine Learning Tools and Techniques: Elsevier Science: New York, NY, USA, 2011.
73. Farquad, M.A.H.; Ravi, V.; Raju, S.B. Churn prediction using comprehensible support vector machine: An analytical CRM application. Appl. Soft Comput. 2014, 19, 31–40. [CrossRef]
74. Pinto, T.; Sousa, T.M.; Praça, I.; Vale, Z.; Morais, H. Support Vector Machines for decision support in electricity markets’ strategic bidding. Neurocomputing 2016, 172, 438–445. [CrossRef]
75. Abdel-Zaher, A.M.; Eldeib, A.M. Breast cancer classification using deep belief networks. Exp. Syst. Appl. 2016, 46, 139–144. [CrossRef]
76. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. Adv. Neural Inf. Proc. Syst. 2007, 19, 153.
77. Bengio, Y. Learning Deep Architectures for AI. Found. Trends Mach. Learn. 2009, 2, 1–127. [CrossRef]
78. Wang, G.; Sun, J.S.; Ma, J.; Xu, K.Q.; Gu, J.B. Sentiment classification: The contribution of ensemble learning. Decis. Support Syst. 2014, 57, 77–93. [CrossRef]
79. Zhou, Z.-H. Ensemble Methods: Foundations and Algorithms, 1th ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2012; p. 236.
80. Abbasi, A.; Chen, H. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. ACM Trans. Inform. Syst. 2008, 26, 7. [CrossRef]
81. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. IEEE Trans. Syst. Man Cybern. C 2012, 42, 463–484. [CrossRef]
82. Diez-Pastor, J.F.; Rodriguez, J.J.; García-Osorio, C.I.; Kuncheva, L.I. Diversity techniques improve the performance of the best imbalance learning ensembles. Inform. Sci. 2015, 325, 98–117. [CrossRef]
83. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 2002, 16, 321–357. [CrossRef]
84. WHO. Suicide Mortality Rate: Data Tables; World Health Statistics: Geneva, Switzerland, 2017.
85. Suh, J.H.; Park, S.C. Service-oriented Technology Roadmap (SoTRM) using patent map for R&D strategy of service industry. Exp. Syst. Appl. 2009, 36, 6754–6772.
86. Lombard, M.; Snyder-Duch, J.; Bracken, C.C. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. Hum. Commun. Res. 2002, 28, 587–604. [CrossRef]

© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).