Hindawi
Security and Communication Networks
Volume 2018, Article ID 7243296, 8 pages
https://doi.org/10.1155/2018/7243296

Research Article

Detecting Potential Insider Threat: Analyzing Insiders' Sentiment Exposed in Social Media

Won Park, Youngin You, and Kyungho Lee

Institute of Cyber Security & Privacy, Korea University, Seoul 02841, Republic of Korea

Correspondence should be addressed to Kyungho Lee; kevinlee@korea.ac.kr

Received 9 March 2018; Revised 2 June 2018; Accepted 26 June 2018; Published 18 July 2018

Academic Editor: Ilsun You

In the era of Internet of Things (IoT), impact of social media is increasing gradually. With the huge progress in the IoT device, insider threat is becoming much more dangerous. Trying to find what kind of people are in high risk for the organization, about one million of tweets were analyzed by sentiment analysis methodology. Dataset made by the web service "Sentiment140" was used to find possible malicious insider. Based on the analysis of the sentiment level, users with negative sentiments were classified by the criteria and then selected as possible malicious insiders according to the threat level. Machine learning algorithms in the open-sourced machine learning software "Weka (Waikato Environment for Knowledge Analysis)" were used to find the possible malicious insider. Decision Tree had the highest accuracy among supervised learning algorithms and K-Means had the highest accuracy among unsupervised learning. In addition, we extract the frequently used words from the topic modeling technique and then verified the analysis results by matching them to the information security compliance elements. These findings can contribute to achieve higher detection accuracy by combining individual's characteristics to the previous studies such as analyzing system behavior.

1. Introduction

We are living in the world of IoT. Everything from a portable device to a home device is being connected to the Internet. A statistic research predicted that above 30 billion devices would be connected to the Internet in 2020 [1]. Furthermore, the number of social media users all over the world is predicted above three billion in 2021 [2]. Social networking is a type of communication that has continued since the creation of mankind, although the forms are different. Also, the current form of social media is in line with the development of Internet and IoT. As the IoT paradigm expands, interaction with social media becomes more active than in the mobile environment, which is expanding to the concept of SIoT (Social Internet of Things) [3]. Also, social media can be used as an interaction tool with the IoT devices [4]. In this way, social media research based on IoT is going on in a variety of ways. As the IoT device develops, the insider threat of the organization is getting bigger. IoT technology expands to become a part of people's lives and provides convenience to people, but efforts and costs for information security are increasing as the number of connection points with existing systems grows [5] because cyber threat such as Denial of Service (DOS) and information leakage can be caused easily by normal or malicious insiders' mobile devices. Also, many studies about analyzing system behavior to detect insider threats have been conducted, but relatively few researches about analyzing emotions that affect individual behavior have been conducted. In order to detect threats, social media analysis is needed, which is one of the media in which individual tendencies are well presented.

In this research, we aim to find the possible malicious insider to prevent the insider threat. A research indicates that many organizations are in face of insider threats. According to the 2018 Insider Threat Report by Cybersecurity Insiders and Crowd Research Partners, 90 percent of organizations felt they were vulnerable to insider threats, and 56 percent answered that regular employees among insiders were the biggest threat [6]. In this situation, many organizations focus on reinforcing internal system. They tend to think that internal threats, as well as external threats, can be somewhat defensible if only information security solutions are built

well. However, C. Colwill insists that information security should not lean in only the technological solution. He cited the survey and found that 94% of UK government agencies had data encryption for wireless networks, 97% used internal firewall protection, and 95% used email scans for protect from external threat. On the other hand, 70% of the frauds were caused by insiders, but 90% of the security controls and monitoring are concentrated on responding to external threats [7]. In this situation, N. Safa, R. Solms, and S. Furnell [8] have an idea that organizations ignoring individual security problem are possible to fail. Furthermore, they indicated that Information Security Organizational Policies and Procedures (ISOP) can improve employee's information security level. The basic theory of their research is Social Bound Theory (SBT) and Involvement Theory. SBT is the theory that the more individuals interact with others, the less likely they are to behave abnormally. In other words, it can be judged that a person who actively exchanges in an organization is unlikely to be a malicious insider and will not cause an insider threat. Involvement theory is that people who actively participate in certain activities are less likely to behave abnormally. There is some overlap with SBT, but it is assumed that the more active people are, the less likely they are to cause insider threats. In short, we are focusing on lowering the risk from insider threats especially by analyzing social media. Through this research, we aim to contribute to the detection of insider threats by combining the sentiment analysis of individuals with the concept of information security compliance. In addition, we expect to be able to detect malicious insiders with higher accuracy through combination with previous studies analyzing system behavior.

2. Related Works

There are many researches about the usage of social media to prevent threats in both real and cyber space. In terms of the insider threat, there are two kinds of research areas: technological and psychological. The first is researches through technological analysis. M. Salem, S. Hershkop, and S. Stolfo [9] presented a methodology for detecting malicious insiders through host and network-based user profiling. Host-based user profiling has provided a way to identify users in environments such as UNIX, Windows, and the web. Network-based user profiling provides a way to identify users by analyzing network traffic such as HTTP, SMB, SMTP, and FTP. Also, they set the criteria for detecting a "masquerader" and an "internal traitor" through two analysis methods. Masquerader is a type of malicious insider who steals and impersonates legal insider. Internal traitor is another type of malicious insider who has been accepted to access the system. A. Harilal et al. [10] simulated malicious behavior in an organization's system. They collected them from seven types of data source: mouse, key strokes, host activities, network traffic, and so on. Through several steps of periods, they collected benign/malicious users' behavior dataset. This dataset gives a chance to utilize a research in malicious insider's behavior in the organization's system. These analytical methods have the merit that they enable the formal analysis, but it has a disadvantage that it is difficult to reflect the tendency of the individual. Therefore, applying the methodology presented in this study makes it possible to detect malicious insiders more effectively. P. Legg et al. introduced a tool for detecting malicious insider called Corporate Insider Threat Detection (CITD) [11]. They conducted profiling of user and role based on system logs such as logins, portable storage usage, and email transmission. This research showed the possibility of adapting decision making process and decreasing the false positive rate when a system detects malicious behavior. Lastly, A. Tuor et al. analyzed CERT Insider Threat Dataset v6.2 of Carnegie Mellon University [12]. They found that the Neural Network model has higher performance than the Principal Component Analysis (PCA), Support Vector Machine (SVM), and Isolation Forest based on anomaly detection. They modeled system's normal behavior and set abnormal behavior as criteria of anomaly. M. Bishop et al. approached as the aspect of process [13]. They analyzed insider's behavior into the process model. In the process model, specific behaviors are broken into the subprocesses. Researches in this part contributed to establish a theory to detect malicious insiders based on the system behavior. However, it is important to understand the insiders' emotional state in order to analyze why they do malicious acts.

In this context, there are researches through psychological analysis. M. Kandias et al. [14] proved that psychological characteristics of negative attitude are closely related to insider attacker's malicious behavior. They compared three kinds of methodologies according to the learning accuracy: machine learning approach, dictionary-based approach, and flat data classification. They used data of well-known streaming service YouTube's comments subscribers, and video views, and so on. Among these methodologies, machine learning approach showed the highest accuracy because many words are learned by learning model. On the other hand, words had to be searched entire list to find what kind of words is the same as the list in the dictionary-based approach. Furthermore, flat data classification is harder to find the same words. V. Marivate and P. Moiloa [15] showed how to detect crime incidents by gathering and analyzing social media data. To prove this, they checked how much is corresponded with the real incident and the estimated event. Many studies have focused on the analysis of behavior of both system and user in an organization. L. L. Ko et al. approached to the insider threat in various areas such as cyber behaviors and communication behaviors [16]. This research has the meaning for analyzing insider threats in perspective of not only the cyber behavior and communication, but also the biometric and psychosocial behavior. Furthermore, they suggested future directions of the insider threat related to their perspectives. B. A. Alahmadi, P. A. Legg, and R. C. Nurse approached in perspective of OCEAN personality analysis [17]. They proved the correlation between Internet usage log and personal trait. Through this research, they were able to identify psychological characteristics by trends in the Internet usage logs and ultimately use them to detect potential malicious insiders. Together with system behavior analysis, researches about psychological aspects contribute to overcome the limitations of technological approaches by understanding the emotions of insiders.

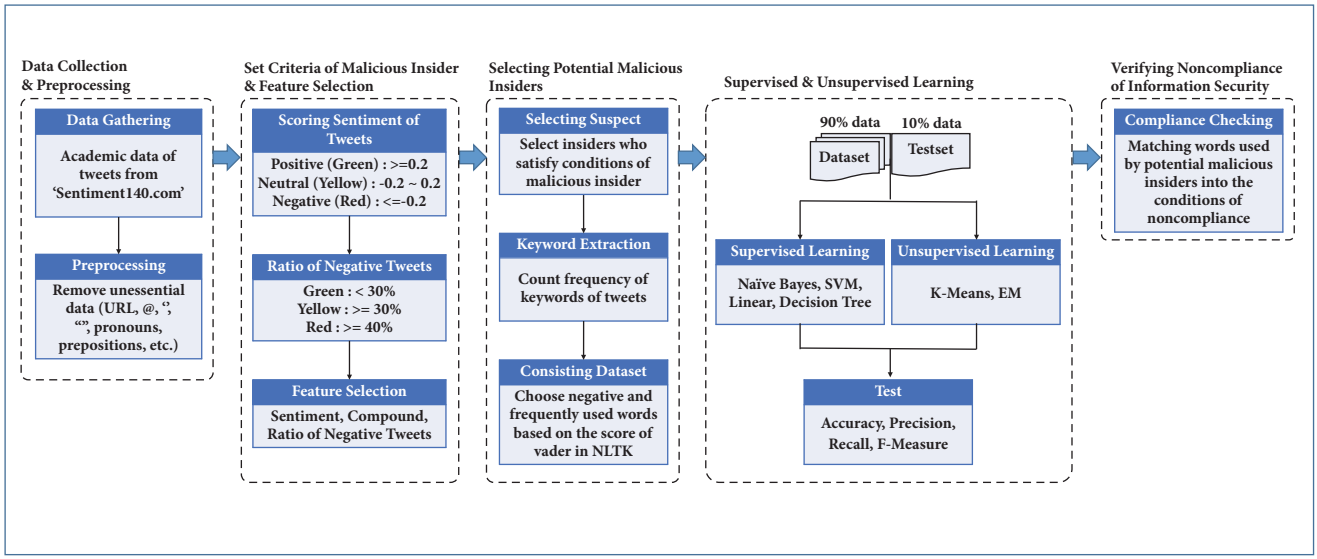

Figure 1: Overall process of the research.

3. Research Methods

In this section, we will introduce some methods which were adapted to analyze data. It is including description of the dataset, methodologies such as machine learning, sentiment analysis, topic modeling, and information security compliance. Overall process of this research is descripted in Figure 1.

3.1. Dataset. For this research's purpose to find a possibility of an insider threat, we needed to get data which are including many people's thinking. Training data from the web service "Sentiment140" is containing almost 1.6 million of tweets and the user who wrote tweets [18]. Among many social media platforms, Twitter is widely used because it can freely analyze using open API. Even though they already having service of sentiment analysis, we analyzed the sentiment of tweets again using open-sourced Python API called NLTK (Natural Language ToolKit). Originally, dataset of Sentiment140 has three levels of sentiment: negative (0), neutral (2), and positive (4). It is similar to the NLTK in terms of the kinds of sentiment. However, it is hard to understand the level of sentiment because it has only fixed sentiment score. To overcome the limitation, NLTK was used to get sentiment level more specifically (from -1 to 1).

3.2. Sentiment Analysis. NLTK is a well-known sentiment analysis tool in python language. It is a suit of program modules, data sets, and tutorials supporting research and teaching in computational linguistics and natural language processing [19]. Each word has an evaluated score used as criteria called "vader". NLTK calculate score to determine how positive/negative is the sentence's or document's sentiment level. We operated NLTK to overall tweets and got each word's sentiment level including overall, positive, negative, and neutral.

3.3. Topic Modeling. It is a statistical model used in natural language processing (NLP) to discover a subject of documents. Latent Dirichlet Allocation (LDA) is the basic topic model and is called a probabilistic model. It also has a generative process including hidden variables and defines a probability distribution [20]. It assumes that the words related to specific topic would appear more often than the other words. At first, all of sentences were broken into the words in corpus and all unnecessary words were deleted such as pronouns, prepositions, and articles.

3.4. Machine Learning. We should find the pattern of dangerous behavior of the insider threat. Both supervised and unsupervised learning algorithms were used because there are different advantages. In terms of the supervised learning, it has an advantage that we can give an answer of who has the possibility of doing dangerous behavior into the dataset. However, data labelling requires highly expensive, time-consuming works [21]. So, too much time could be consumed because almost 1 million of data were used. In the aspect of efficiency, unsupervised learning can be advantageous. Open-sourced machine learning software called "Weka" was used to compare the accuracy among several algorithms. The following are supervised/unsupervised learning algorithms.

3.4.1. Supervised Learning. Supervised Learning is useful when an exact answer is given. In this research, we labelled possible malicious insider. There are four supervised learning algorithms used in classification. Naïve Bayes is a generative model by using assumption to reduce the parameters to learn [22]. When there are n features, the number of parameters must be minimized since the example to be learned increases exponentially. Finally, it estimates the Naïve Bayes classifier which is most likely to be used as features. Support Vector

Machine (SVM) is a suitable algorithm to adapt text classification. In particular, it is being widely used to classify news stories and search result. T. Joachims [23] suggested the several advantages to apply the text classification and showed that SVM is the most accurate algorithm among other classification algorithms. Linear algorithm is normally used algorithm for a statistical case like predicting baby's body weight [22]. It has a loss function which calculates how accurately it predicts. So, Empirical Risk Minimization (ERM) is widely used to minimize the loss value. Lastly, in the Decision Tree algorithm, root makes branches from learned data and classifies each data through conditions of leaves. In terms of the NLP, this algorithm can be utilized to categorizing document, parts of speeches [24].

3.4.2. Unsupervised Learning. Unsupervised Learning is useful when an exact answer is not given. K-Means is a well-known algorithm so that it can adapt many unsupervised learning cases. Its basic principle is simple. First, select k centroids. Second, calculate cost value (distance) between each centroid and element. Lastly, adjust centroids' location and repeat these processes. Expectation-Maximization (EM) focuses on reducing log likelihood and its processes are executed repeatedly until mixture model is close to the current log likelihood [25]. There are two steps to deal with the data: E-step and M-step. E-step is a computing process to get expected value using currently predicted parameter and discovered data. M-step is a process of determining the value of the parameter using predicted data called "imputed" data. In this process, it assumes that the data of the expectation step is actually measured data [26]. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is another type of clustering algorithm. It has two kinds of parameters: Eps and MinPts. Eps means a distance between the core point and a point in a cluster. MinPts means the number of points. Cluster can be made when there are MinPts of points. It uses heuristic method to determine parameters and finds "thinnest" cluster that has the limit value of the maximum k-dist value [27].

3.5. Information Security Compliance. In the current IT environment, insider can cause massive harm and we cannot prevent or mitigate from the damage of malicious insiders. So, human factors should be considered to overcome the limit of technology-based information security compliance [7]. We reorganized research table of the information security compliance of [8] and matched each characteristic and used words. Through this process, we demonstrated that possible malicious insiders did not match the conditions of compliance of the information security.

4. Research Results

There are five steps to estimate possible malicious insider by analyzing collected tweets. After preprocessing, overall insiders' sentiment levels were scored and selected possible malicious insiders. Dataset was consisted based on the result of topic modeling. Machine learning algorithms have been used to classify and cluster possible malicious insiders. Finally, selected insiders who are in red group of possibility were adapted to the condition of noncompliance of information security and compared whether he/she corresponds to the condition of noncompliance. Similarly, W. Park, Y. You, and K. Lee [28] researched correlation between tweets and actual behavior in the real world. 4,000 Tweets of 2016 U.S. presidential election nominees (Donald Trump and Hillary Clinton) were crawled. After sentiment analysis process, daily events were correlated with the negative tweets based on daily average sentiment score. Through the machine learning process, this research proved that individual's social media reflects the real world's situation.

4.1. Preprocessing. The dataset of "Sentiment140" consisted of .csv file. It has user ID, date, tweets, and sentiment of tweet in the file. Some tweets have mathematical character "=". When this character is at the front of the sentence, excel program perceives it as a function. So, we had to remove "=" at the front of sentence. Also, we removed unessential words including the parts of sentence such as pronouns, prepositions, and articles. Through this process, we could get the groups of words and were erased by the topic modeling.

4.2. Analyzing Insiders' Overall Sentiment. At first, sentiment level was scored through the sentiment analysis of the whole data. After that, the average of the sentiment score and the ratio of negative tweets are calculated. Figures 2 and 3 show a part of analysis result. Figure 2 is about each user's sentiment score and Figure 3 is about the ratio of negative tweets. Each figure has a green line which means the criteria of the baseline of the negative sentiment. In this dataset, we classified 400 people depending on the criteria. Most insiders were in the green level, but 1 user (0.25%) was in the red level and 36 users (9%) were in the yellow level.

4.3. Selecting Possible Malicious Insider. In order to find a person who is possible to be a potential threat, we set criteria to classify people in three levels by sentiment score and the ratio of negative tweets. Table 1 shows how to classify people by the threat level. The degree of insider threat was divided into three stages: green, yellow, and red. Each level is assigned if both the sentiment score and the ratio of negative tweets are satisfied. The baseline for judging negative sentiment was set to -0.2 based on [28] and it was converted from [29]. Also, the ratio of negative tweets was set 40%. According to M. Losada and E. Heaphy [30], they separated three types of teams by efficiency: high, medium, and low-performance. In particular, in case of low-performance teams, ratio of positivity and negativity (P/N) is 0.363. Additionally, E. Ferrara and Z. Yang suggested ratio of the neutral messages in the social media [31]. They measured about 45% of messages had neutral emotion. Table 1 shows the criteria which we made to detect possible malicious insider through these researches. These criteria applied to classify the degree of insider threat, and the system administrator can flexibly respond when it comes to the system management by further

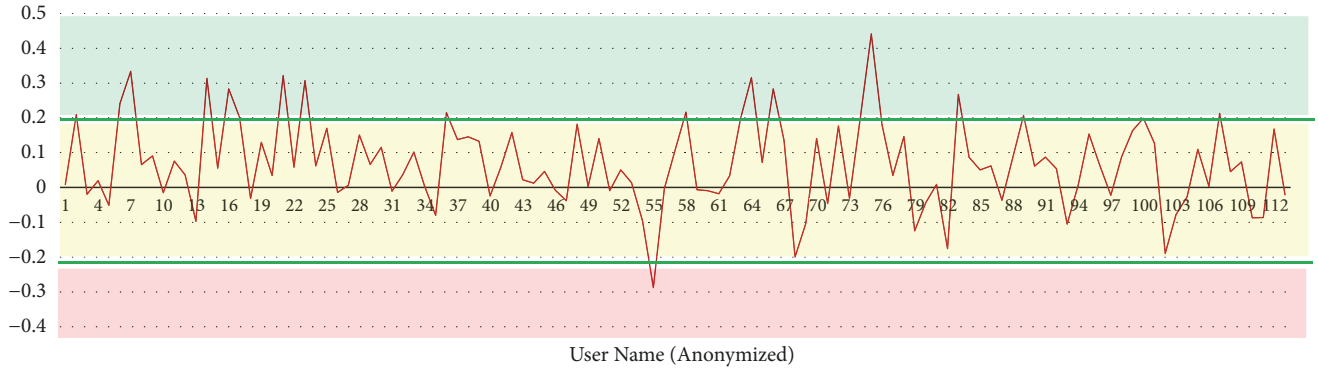

Figure 2: Example of the ratio of the negative tweets.

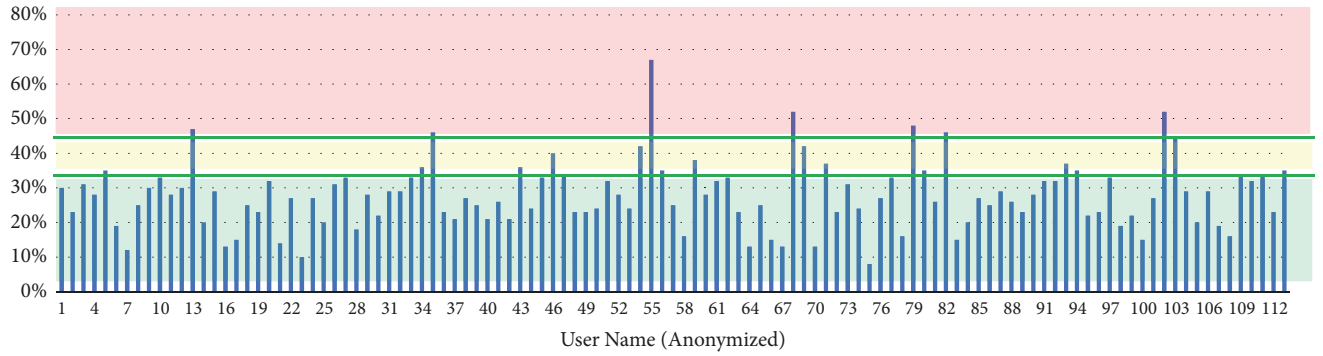

Figure 3: Example of the result of insiders' sentiment analysis.

Table 1: Three levels in classifying potential threats.

Level (Threat Possibility)	Sentiment Score (s)	Ratio of Negative Tweets (r)
Green (Low)	s >= 0.2	r < 35%
Yellow (Medium)	-0.2 < s < 0.2	r >= 35%
Red (High)	s <= -0.2	r >= 45%

Table 2: Comparing possible malicious suspects.

User (Anonymized)	Sentiment Score	Sentiment Level	Ratio of Negative Tweets
A	-0.2873	Red	67%
B	-0.1994	Yellow	53%
C	-0.1892	Yellow	52%
D	0.2409	Green	19%
E	0.2161	Green	16%

subdividing the criteria from the previous research. From these criteria, Table 2 shows five people of each level.

4.4. Machine Learning. As a way to detect malicious insider, Machine Learning methodology was used. Both supervised and unsupervised learning algorithms were adapted in this research and the result of learning is descripted in Sections 4.4.2 and 4.4.3. This process has a meaning as a way to identify possible malicious insiders and improve the accuracy. Topic Modeling was used to rank according to how often the words in tweets were used. And then, we listed 100 of negative words. Table 3 shows the example of dataset which was adapted to the research. We focused negative words because we assumed that people's negative thinking related to the threat would be presented by the words. To make a dataset, negative words were selected as features. Each of the features (word) and the number of how many times they were used were included in the dataset.

4.4.1. Feature Selection. There are so many words which consist all of tweets. Among them, we chose 100 of negative words by frequency. After that, we composed a dataset and checked if the word is in the sentence.

4.4.2. Supervised Learning. Several kinds of supervised learning algorithms were adjusted to analyze the accuracy of learning. Table 4 shows the result of supervised learning. Decision Tree had the highest accuracy of learning and is followed by SVM, linear, and Naïve Bayes. A kind of probabilistic model Naïve Bayes assumes that the population follows Gaussian or polynomial distributions, but the data in the research is not exactly of a distribution type, so it can have relatively low accuracy. The SVM can produce high accuracy when it classifies two categories by creating a nonprobabilistic

Table 3: Example of the dataset.

User	Sentiment Score	Ratio of Negative Tweets	Label	word1 (sad)	word2 (miss)	word3 (bad)	...	word100 (blocked)
A	-0.2873	0.67	Yes (1)	1	0	1	...	1
B	-0.1994	0.52	Yes (1)	0	0	1	...	0
C	-0.1892	0.52	Yes (1)	0	0	0	...	1
D	0.2409	0.19	No (0)	1	0	0	...	0
E	0.2161	0.16	No (0)	0	1	0	...	0

Table 4: Result of Supervised Learning.

Algorithm	Accuracy	Precision	Recall	F-Measure
Naïve Bayes	96.2%	1.000	0.962	0.980
SVM	99.5%	0.991	0.996	0.993
Linear	99.5%	0.994	0.995	0.994
Decision Tree	99.7%	0.998	0.997	0.996

Table 5: Result of Unsupervised Learning.

Algorithm	Accuracy	Incorrectness
K-Means	95.6%	4.4%
EM	83%	17%
DBSCAN	90.7%	9.3%

Table 6: Frequently used words.

User (Anonymized)	Sentiment	Words
A	Positive	can, packing, good, working, ...
	Negative	miss, bad, hate, wrong, hell, ...
B	Positive	like, want, awesome, hope, fine, ...
	Negative	damn, bad, sad, hell, dying, poor, ...

binary linear classification model, and the linear algorithm has similar characteristics. The Decision Tree algorithm has the property of iterating until a new prediction is no longer added because the process is iteratively recursive and can have a relatively high accuracy.

4.4.3. Unsupervised Learning. Several kinds of unsupervised learning algorithms were adjusted to analyze the accuracy of learning. Table 5 shows the result of unsupervised learning. K-Means had higher accuracy than EM and DBSCAN. Unsupervised learning result was relatively inaccurate because learning process occurs without information about who the malicious insider was. The K-Means algorithm can improve the accuracy of detection by minimizing the dispersion of the cluster and distance difference while forming a certain cluster. The EM algorithm can be used to obtain the maximum likelihood when the statistical model cannot be solved correctly, but it may not converge to the maximum possible degree. DBSCAN basically forms clusters based on eps and MinPts, but the dataset has some parts that do not fit well into the conditions.

4.5. Get Frequently Used Words. Topic modeling was used and ranked into the frequency of how often the words were used. We focused on negative words because we assumed that people's negative thinking related to the threat would be presented in the words. All the words in Table 6 were extracted by topic modeling from each user's tweets. Basically, words were classified to positive and negative based on NLTK's sentiment score database (vader_lexicon.txt). There are a lot of words and each score was evaluated by several people.

4.6. Possible Malicious Insider Selection. All of tweets are analyzed and classified within the criteria to detect the highly possible malicious insiders. Also, we only selected people who wrote at least 45 tweets. The words they used were matched to condition of the compliance. Finally, the most possible malicious insiders were selected. Table 7 shows the example of verifying insider threat based on the concept of information security compliance.

5. Conclusions

Social media can help people around the world communicate freely. Its role is also diversifying as the Internet environment has been expanding to the mobile and IoT-based. At the same time, we can use the social media as useful analysis media to detect the potential insider threat. In this paper, it has been shown that it is efficient to analyze individual tendencies to detect possible malicious insider. To do this, sentiment analysis was conducted on the collected tweets, and we classified the users by the level of threat according to criteria. And then, the classified possible malicious insiders were verified by performing information security compliance matching process. Machine learning algorithms were applied to detect possible malicious insiders. Decision Tree algorithm could detect possible malicious insiders with the highest accuracy of 99.7%. In addition, user A was expected not to meet information security compliance as using the words such as miss, hate, and wrong. In this way, a methodology has been proposed to prevent damage to the organization's information systems. Meanwhile, to prevent conflicts with privacy issues, it is necessary to sufficiently inform and consent internally that this analysis is merely a means for detecting insider threats and not for privacy violations. This

Table 7: Verifying malicious insider through the concept of information security compliance.

Construct	Description	Yes / No	Used Words
Spreading Knowledge of Information Security	He / She doesn't want to give knowledge of information security to the other people.	✔	hell, bad
	Others don't help to improve his / her knowledge of information security.		
	He / She doesn't like to join a seminar related to information security.		
Cooperating with Others	Cooperation is unhelpful to decrease the risk of our organization.	✔	wrong, badly
	He / She won't contribute to improve our organization through cooperation.		
Learning Information Security	He / She doesn't want to learn new things.	✔	old, hate
	Accepting new things are not helpful to him / her.		
Practicing Information Security	His / Her experience can't affect to organization positively.		
	Experiences don't motivate him / her.		
Attachment	He / She doesn't like to communicate others.	✔	miss, hate,
	He / She doesn't tend to obey the rules.	✔	damn, wrong
Commitment	He / She is careless about keeping valuable things safely.	✔	forgot
	He / She doesn't try to follow up the newest policies.	✔	weird

paper contributes to the analysis of data on social media to show that the criteria for detecting insider threats are based on the sentiment level and the ratio of negative emotions, and it can be verified based on the concept of information security compliance. Recent insider threats are not just the actions of the system suspension or information leakage, but it is very important to prevent the attack, because of the organization's reputation, customer compensation, and so on. Above all, to improve the level of information protection of the organization, it is necessary that not only the information protection person but also the management's active interest and efforts are put together. In the future, we will conduct research on methodology that can combine analysis of system behavior and individual's sentiment analysis to detect insider threat.

Data Availability

The .csv formatted data used to support the findings of this study may be released upon application to the "Sentiment140.com", who can be contacted at http://help.sentiment140.com/for-students/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2015-0-00403) supervised by the IITP (Institute for Information and communications Technology Promotion).

References

[1] Statista, Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions), 2016.

[2] Statista, Number of social media users worldwide from 2010 to 2021 (in billions), 2017.

[3] A. M. Ortiz, D. Hussein, S. Park, S. N. Han, and N. Crespi, "The cluster between internet of things and social networks: review and research challenges," *IEEE Internet of Things Journal*, vol. 1, no. 3, pp. 206–215, 2014.

[4] M. Kranz, L. Roalter, and F. Michahelles, "Things that twitter: social networks and the internet of things, What can Internet Things do Citiz. Work," in *Proceedings of the 8th International Conference on Pervasive Computing (Pervasive '10)*, pp. 1–10, 2010.

[5] X. Liu, M. Zhao, S. Li, F. Zhang, and W. Trappe, "A security framework for the internet of things in the future internet architecture," *Future Internet*, vol. 9, no. 3, 2017.

[6] Cybersecurity Insiders and Crowd Research Partners, Insider threat 2018, 2017.

[7] C. Colwill, "Human factors in information security: the insider threat—who can you trust these days?" *Information Security Technical Report*, vol. 14, no. 4, pp. 186–196, 2009.

[8] N. S. Safa, R. V. Solms, and S. Furnell, "Information security policy compliance model in organizations," *Computers and Security*, vol. 56, pp. 1–13, 2016.

[9] M. B. Salem, S. Hershkop, and S. J. Stolfo, "A survey of insider attack detection research," *Advances in Information Security*, vol. 39, pp. 69-70, 2008.

[10] A. Harilal, F. Toffalini, J. Castellanos, J. Guarnizo, I. Homoliak, and M. Ochoa, "TWOS: a dataset of malicious insider threat behavior based on a gamified competition," in *Proceedings of the 2017 International Workshop on Managing Insider Security Threats*, pp. 45–56, ACM, 2017.

[11] P. A. Legg, O. Buckley, M. Goldsmith, and S. Creese, "Caught in the act of an insider attack: detection and assessment of insider

threat," in *2015 IEEE International Symposium on. IEEE*, pp. 1–6, 2015.

[12] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," https://arxiv.org/abs/1710.00811.

[13] M. Bishop, H. M. Conboy, . Huong Phan et al., "Insider Threat Identification by Process Analysis," in *Proceedings of the 2014 IEEE Security and Privacy Workshops (SPW)*, pp. 251–264, San Jose, Calif, USA, May 2014.

[14] M. Kandias, V. Stavrou, N. Bozovic, and D. Gritzalis, "Proactive insider threat detection through social media: The YouTube case," in *Proceedings of the 1st ACM Workshop on Language Support for Privacy-Enhancing Technologies, PETShop 2013 - Co-located with the 20th ACM Conference on Computer and Communications Security, CCS 2013*, pp. 261–266, Germany, November 2013.

[15] V. Marivate and P. Moiloa, "Catching crime: Detection of public safety incidents using social media," in *Proceedings of the 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference, PRASA-RobMech 2016*, South Africa, December 2016.

[16] L. L. Ko, D. M. Divakaran, Y. S. Liau, and V. L. L. Thing, "Insider threat detection and its future directions," *International Journal of Security and Networks*, vol. 12, no. 3, pp. 168–187, 2017.

[17] B. A. Alahmadi, P. A. Legg, and J. R. Nurse, "Using Internet Activity Profiling for Insider-threat Detection," in *Proceedings of the 12th Special Session on Security in Information Systems*, pp. 709–720, Barcelona, Spain, April 2015.

[18] "Dataset of Sentiment140," http://help.sentiment140.com/for-students/.

[19] E. Loper and S. Bird, "NLTK: The Natural Language Toolkit," in *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pp. 1–4, 2004.

[20] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[21] A. Dundar, J. Jin, and E. Culurciello, "Convolutional Clustering for Unsupervised Learning," pp. 1-11, 2015.

[22] S. Ben-David and S. Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*, 2014.

[23] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 1398, pp. 137–142, 1998.

[24] S. Schrauwen, *Machine Learning Approaches to Sentiment Analysis Using the Dutch Netlog Corpus*, 2010.

[25] P. Bradley, U. Fayyad, and C. Reina, "Scaling EM (Expectation-Maximization) Clustering to Large Databases," *Microsoft Research*, pp. 1–25.

[26] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[27] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD '96)*, vol. 96, pp. 226–231, 1996.

[28] W. Park, Y. You, and K. Lee, "Twitter sentiment analysis using machine learning," *Research Briefs on Information & Communication Technology Evolution*, http://rbisyou.wixsite.com/rebicte/volume-3-2017, 2017.

[29] S. Feng, J. S. Kang, P. Kuznetsova, and Y. Choi, "Connotation lexicon: a dash of sentiment beneath the surface meaning," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 1774–1784, 2005.

[30] M. Losada and E. Heaphy, "The Role of Positivity and Connectivity in the Performance of Business Teams: A Nonlinear Dynamics Model," *American Behavioral Scientist*, vol. 47, no. 6, pp. 740–765, 2004.

[31] E. Ferrara and Z. Yang, "Measuring emotional contagion in social media," *PLoS ONE*, vol. 10, no. 11, Article ID e0142390, 2015.