big data and
cognitive computing

MDPI

Article

Analysis of Information Security News Content and Abnormal Returns of Enterprises

Chia-Ching Hung

Department of Information Management, National Chung Cheng University, Chiayi County 62102, Taiwan; cch@mis.ccu.edu.tw

Received: 27 March 2019; Accepted: 24 April 2019; Published: 27 April 2019

Abstract: As information technologies and the Internet have rapidly evolved, businesses have begun to use them to improve communication efficiency within and outside the organization. However, applications of information technologies are accompanied by information delivery, personal data protection, and information security problems. There are potential risks inherent in any application of information technologies. Moreover, with the improvement of networking and computing capabilities, the impact of attacks from hackers and malicious software has also increased. A breach or leakage of important corporate data may not only damage the firm's image but also sabotage the firm's operation, resulting in financial losses. In this study, the content of information security news reports was analyzed in an attempt to clarify the association between information security news and corporate stock prices. Methods including decision trees, support vector machines (SVMs), and random forests were used to explore the associations of news related variables with abnormal returns. Results indicate that the news source and the presence of negative words in the news have an impact on abnormal returns.

Keywords: information security news; market value; abnormal return; decision tree; support vector machine; random forest

---

1. Introduction

In recent years, with the increasing prevalence of the Internet and cloud infrastructure, more and more businesses have begun to utilize information systems to improve the efficiency of their internal and external operations. As a result, many problems surrounding information technology management have arisen. When a company's operational data can be quickly diffused within the organization and even across organizations, ensuring the integrity, confidentiality, and accessibility of the data is an issue that the company should seriously deal with while reaping the benefit of high efficiency [1].

In addition, as network transmission capacity and computer performance have continuously improved, hackers are able to steal important corporate secrets or consumers' personal information within a short time. Attacks from hackers and malicious software, and the resulting damage, are steadily growing, making assurance of information security more important than ever. In several information security events, such as the recent data leak suffered by Cathay Pacific Airways, where 9.4 million passengers were affected, all the affected companies suffered serious damage to their corporate image, a decline of their customers' confidence in the firm's website and system, and a tremendous financial loss.

An information security vulnerability can cause losses in a wide range of aspects. In addition to a reduction in revenue, the company may suffer intangible losses that are even greater. Take the event where customers' credit card data are stolen due to inadequate security protection as an example. The data breach will result in a collapse of customer confidence and trust. The loss of this kind of intangible asset is usually greater than the loss in business operations.

Due to cost considerations, firms usually pay little attention to or ignore the importance of information security. As a result, all kinds of security events still periodically occur. Therefore, by analyzing news coverage of information security events and stock prices, this study attempts to highlight the potential effect of an information security event on stock prices, providing evidence that reminds firms to attach importance to information security and establish a reliable and safe platform for their users.

In addition, in recent years, fake news has been increasingly rampant, and many media have been found to report news with a bias. In the analysis of news content, this study will analyze the use of negative words and certain keywords in articles. The results can be a reference for news media, and can contribute to higher accuracy and impartiality of news reporting.

This study relies on an integrated application of the event study method and various analysis methods to explore the effect of security events on stock prices, and whether the wording and phrasing of news articles also affect stock prices of the reported firm. In the analysis of abnormal returns, this study collects information security events that occurred in recent years and excludes events that may affect the calculation of abnormal returns, such as mergers and acquisitions. After proper samples are selected for analysis, this study calculates if the company involved in the event suffered negative abnormal returns. Later, a content analysis of the news articles will be carried out. By extracting and analyzing keywords and negative words related to information security, this study attempts to examine if the presence of any keywords and negative words in a news report results in abnormal returns.

Finally, through validation of hypotheses and analysis of wording and phrasing of news articles, this study hopes to increase firms' awareness of the security of their information systems. In so doing, it seeks to motivate firms to reduce the loss of profits or damage to corporate image resulting from a security breach or hacker attack, and to consolidate their infrastructure and employee training to enhance the security of their information systems.

In prior research of the association between information security and stock price fluctuation, event study is a common method employed to examine whether news coverage of certain events has an effect on the rise or decline of the company's stock price. The event study method explores investors' responses to news that is just released to the market. The objective is to examine whether the occurrence of a certain event induces an abnormal change in the firm's stock price and further creates abnormal returns (AR).

In this study, an event study analysis, statistical testing methods, and the decision tree approach are used to explore the correlation between information security events and abnormal returns. Later, the source, use of negative words, and other variables of news articles are analyzed using a number of tools, including the decision tree, support vector machine (SVM), and random forest, to identify their respective associations with abnormal returns.

Previous research data sets are dated and have not been subject to news content analysis. In this study, the information security events published in major newspapers in the U.S. between 1 January 2009 and 31 December 2015 are analyzed. From a further analysis, we also found that the influence of information security news is the greatest on the day following the event. Based on the news content extracted features, this study applies decision tree, SVM, and random forest methods to explore the key factors affecting information security events and abnormal returns, as well as the importance of each variable.

This paper consists of five sections. The first section is an introduction of the research background, motivations, objectives, and contributions. The second section reviews related works and their findings. The third section explains the methodology, including the data collection method and the procedure for event study. The fourth section explains the data analysis and experimental results. The fifth section concludes this study and offers suggestions for future researchers.

2. Literature Review

In business administration research, event study is one of the most extensively used methods. This section discusses the relationship between event study and information security breaches and the literature on analysis methods.

2.1. Event Study

In business administration research, event study is one of the most extensively used methods. The purpose of an event study is to assess whether a firm's stock price fluctuates following a major event. So far, this method has been widely applied in research of the effects of news events, including corporate merger, executive officer turnover, and update and maintenance of an information system. Results of these studies show how news coverage or announcement of a major event affects a firm's stock price. There are also a number of studies addressing the association between information security and stock prices.

A study of related events can offer an insight into whether a firm's market price is related to a specific event. Hence, the event study method is frequently applied in empirical research of information technology, financial, accounting, and management issues [2–6]. In addition, the case study method is also applied in research of big data [7], information applications [8,9], and business intelligence systems [10]. Table 1 provides a list of studies probing into factors affecting stock prices.

Table 1. Prior studies on factors affecting stock prices.

Author(s)	Factor(s) Affecting Stock Prices	Major Finding
Carow et al., 2004	Expanding business through a merger or acquisition	When a merger or acquisition is initiated, the firms involved will experience positive abnormal returns.
Arya & Zhang, 2009	Institutional reforms driven by new regulations and corporate social responsibility requirements	Institutional reforms aimed at fulfilling CSR (Corporate Social Responsibility) are viewed positively by investors.
Konchitchki & O'Leary, 2011	Corporate investment in knowledge management systems	Investment in knowledge management systems can create a 1.25% positive abnormal return and bring positive benefits to the future operation of the company.
Rubin & Rubin, 2013	Business Intelligence (BI) systems	Business Intelligence (BI) systems facilitate sharing of information, so they can not only support decision-making but also help reduce the firm's financial risks and stock price volatility.
Son et al., 2014	Investment in cloud computing	Investing in cloud computing brings significantly positive abnormal returns to a company.
H. Songur & J.E. Heavilin, 2017 [11]	Research and development (R&D) investments	Firms proactive in R&D investment tend to have significantly positive abnormal returns.
Song et al., 2017 [12]	Merger and acquisition (M&A)	Generally, M&A creates positive abnormal returns to acquirers.
Modi et al. 2015 [13]	Security breaches and negative abnormal returns	Security breaches that arise from negligence of a third-party front-end service provider may result in negative abnormal returns of the buyer firm.
Leitch & Sherif, 2017 [14]	Twitter mood, CEO succession announcements and stock returns	There is a negative relationship between Twitter sentiment and stock returns. Older CEOs are associated with larger fluctuations of the stock price.

Although applications of the case study method may slightly vary depending on the research area, the procedure is basically the same. First, it is necessary to determine the categories of the news events to be collected, such as corporate merger and information security events. Later, through queries of news databases, the event day of each event can be confirmed, and whether there are other factors having an interaction effect on the event that should be examined. If there is any factor causing a confounding effect, the news event should be excluded. The next step is to define the estimation period and the event window, estimate the abnormal returns, and test the proposed hypotheses using statistical methods. Finally, the testing results are analyzed and discussed.

2.2. Studies of Information Security Breaches

Prior studies of information security events based on an event study approach have addressed issues including information security strategies [15,16], investment in information security [17,18], and the association between top management involvement and information security management [19]. Want et al. mention in their study that news media is an important source of information, and includes newspapers and blogs [6]. When news about a firm's information security vulnerability is reported, the news is very likely to cause short-term and abnormal variations in the amount and the price of stocks that are traded on the market. Gordon et al. mention that with the development of technologies and e-commerce, information security has become particularly important, and businesses have become more willing to allocate more funding to protection of information security and disclosure of related information to the public [20]. Sufficient disclosure of information system security information increases users' understanding of the system and intention to use the system. In addition, when litigation arises from the system's vulnerability, due to sufficient disclosure of related information to users, the litigation cost for the company can be lower. Gordon et al. collect 10-K, 10-KSB, and 20-F annual reports submitted by firms to the U.S. Securities and Exchange Commission (SEC) during the period 2000–2004 [20]. In their analysis of these reports, they adopt the information security-related keywords as listed in Table 2.

In today's society, information security is an imperative issue for businesses across all industries.

Table 2. The keywords used in Gordon et al. [20].

Access Control	Cyber Security	Network Security
Authentication	Cybersecurity	Security Breach
Business Continuity	Denial of Service	Security Expenditure
Computer Breach	Disaster Recovery	Security Incident
Computer Intrusion	Encryption	Security Management
Computer Security	Hacker	Security Measure
Computer System Security	Information Security	Security Monitoring
Computer Virus	Infosec	Intrusion
Cyber Attack		

2.3. Analysis Methods

2.3.1. Decision Tree

The decision tree method is a machine learning approach that can be used for data classification and prediction. Users first train a decision tree based on one or multiple training datasets and use it to analyze or identify the relationships between attributes in the dataset to be processed. It is an important classification technique among data mining techniques. Kositanurit et al. use the decision tree method to develop a theoretical model for examining ERP systems and end-user performance. They suggest that the decision tree method is good for connecting theory and research of information systems. Their findings indicate that system administrators can control specific properties of an ERP system to enhance end-user performance [21].

Wang et al. present an application of text mining and a decision tree in data classification. They use text mining to extract news on information security to explore the relationship between information security risk and stock prices [6]. Text mining has been proven effective in extracting nontrivial patterns and trends in data [22]. Text mining has also been applied in areas such as classification of news events, detection of management fraud, and improvement of customer support based on basic financial data [23,24].

The goal of a decision tree analysis is to analyze and predict the variables in the dataset to generate a tree structure. In this tree structure, each node represents a test of an attribute, each branch represents the outcome of the test, and each end node shows the final outcome of the classification. Therefore, the main crux of a decision tree lies in deciding on the appropriate attributes to be the nodes for the most effective classification of data.

Common types of decision trees include CHAID (Chi-square Automatic Interaction Detector), CART (Classification and regression tree), ID3, C4.5 and C5.0. In this study, the C5.0 algorithm will be adopted. In terms of accuracy and efficiency, C5.0 is superior to the other algorithms.

2.3.2. Support Vector Machine (SVM)

A support vector machine (SVM) is a machine learning-based classification algorithm developed by Vapnik et al. based on statistical learning theories [25].

So far, SVM has been applied in research of issues pertaining to business administration and information systems. For example, Pal et al. use stagewise regression and SVM to analyze the business health of about 200 U.S. firms, including firms of both low and high ratings [26]. Niklis et al. apply SVM to create a credit risk rating model combining the option-based approach and the accounting-based approach [27].

Al-Yaseen et al. state that due to the increasing connectivity between computers, intrusion detection has become a key to network security. They design a model to deal with the real intrusion detection problems in data analysis and classify network data into normal and abnormal behaviors. This model uses SVM and an extreme learning machine to enhance the efficiency of detecting known and unknown attacks. In addition, they also propose a modified K-means algorithm to build high-quality training datasets, which can help improve the performance of classifiers [28].

Khalifi et al. apply SVM in an information retrieval system. When the data type is text, the complex nature of the database will make information retrieval more difficult. In the present, semantic relationships or machine learning techniques can be employed to assist with the selection of results to return. The authors propose a formal model and a new search algorithm. This algorithm can find associations between information items and then use them to structure search results. It incorporates a natural language processing stage and a machine learning model to select relevant results [29].

Based on a powerful mathematical foundation and statistical theories, SVM works as a collaboration between statistics and machine learning. It can handle both linear and nonlinear data. The basic concept of SVM is to classify a dataset consisting of two classes of data by finding an optimal hyperplane that maximizes the distance from the hyperplane to the closest vector in either data group [30] (see Figure 1). Because SVM delivers high performance in classification, it will also be adopted in this paper to classify and analyze the importance of variables of news articles.

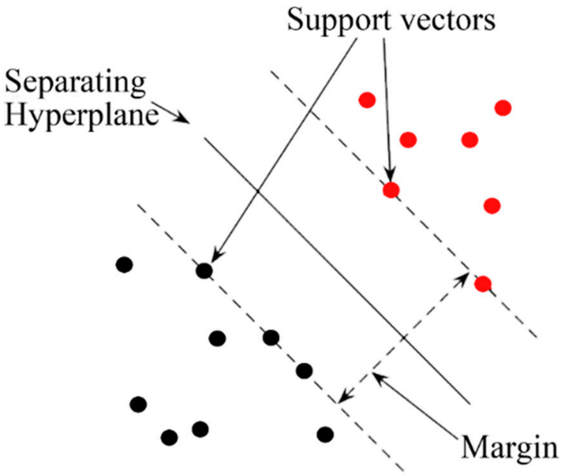


Figure 1. The algorithm of support vector machines.

2.3.3. Random Forest

The basic concept of a random forest is to combine multiple decision trees trained over randomly distributed training data to improve the classification outcome. Its major advantages include that it supports multiple types of data and creates a highly accurate classifier. In addition, it can also estimate missing data and allow missing data or the loss of a portion of data without compromising the classification accuracy. It was proposed by Breiman [31].

Studies applying the random forest technique are numerous. For example, Xie et al. apply a random forest combined with sampling techniques and cost-sensitive learning to customer churn prediction, which is an important issue for the banking sector. The banking sector seeks to satisfy customer needs and retain customers [32]. In Whitrow et al., researchers apply a random forest to credit card fraud detection. They draw on two datasets, each consisting of about 47,000 observations, of which 70% are used for training and 30% for testing. Their results indicate that the random forest is superior to other classification algorithms [33].

Kouzani presents an application of the random forests approach to facial recognition. They use a set of 2414 images represented by 3584 or 3342 variables (gray values of pixels of resized images) and evenly distributed across 38 classes, and find that a random forest consisting of 500 trees and using 100 variables to split a node delivers a better recognition performance than other algorithms [34].

Gupta et al. utilize random forests to predict stock returns. Using U.K. stock data from March 1977 to March 2016, they analyze inequality measures and conditional distribution of stock returns to explore factors causing inequality of income and wealth. Their findings can be a reference for other countries when addressing issues about economic development [35].

As can be seen from the above studies, the random forest is an accurate and efficient algorithm. Therefore, it is also included as one of the analytic algorithms in this study.

This study analyzes incident research methods and information security news events, and integrates the features of news content to explore whether it affects abnormal returns.

3. Research Methods

This section introduces the research process and the process of data collection and extraction, the method of calculating abnormal returns, and analysis of the correlation between news content and abnormal returns.

3.1. Research Procedure

This study begins with collection of news articles about information security events, then analyzes the news articles, and finally performs an event study analysis to identify the correlations. The research procedure is as illustrated in Figures 2 and 3 [36].

3.2. Data Collection and Selection

Most prior studies of security-related events use data mainly from a period between 1998 and 2012. However, due to the rapid advancement of information technologies, the attack types and security issues are much different in the present day, and so are market reactions to security events. Hence, this study collected news on information security events that occurred during the period 2009–2015 for an event study [36].

The news reports on information security events were collected from the Factiva database using the keywords adopted in prior studies [6,37] as well as the keywords mentioned in ISO 27001 standards for information security management systems. Table 3 provides a list of all the keywords used in this study [36].

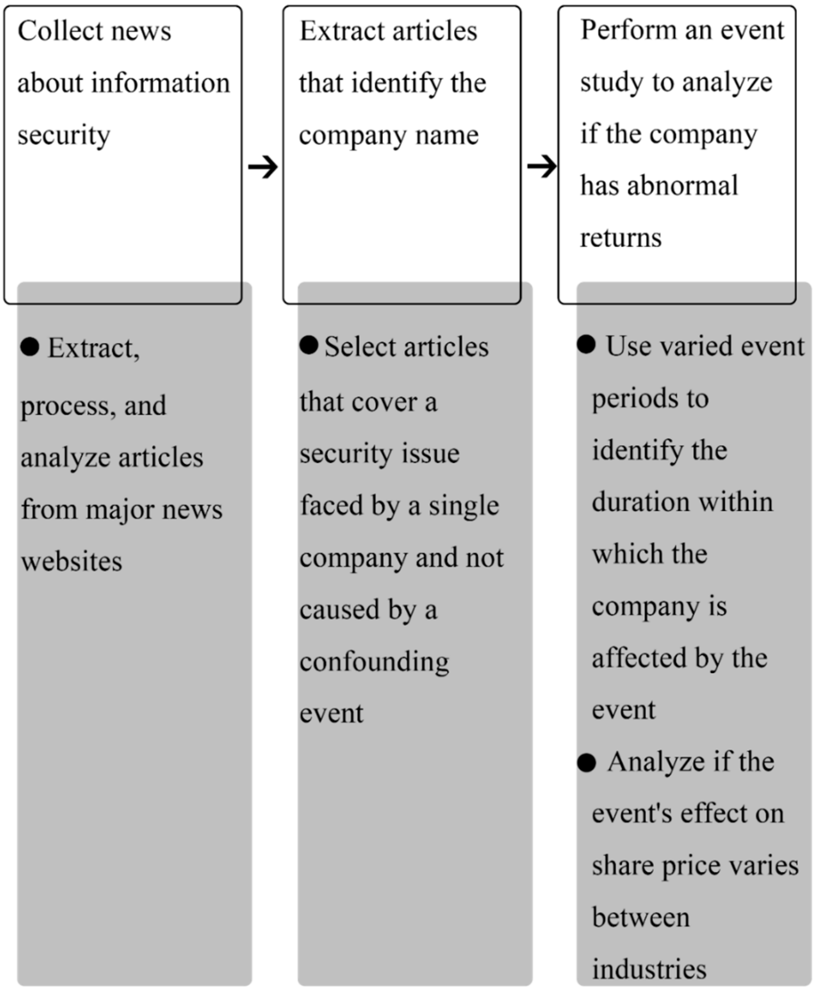

Collect news about information security
● Extract, process, and analyze articles from major news websites
Extract articles that identify the company name
● Select articles that cover a security issue faced by a single company and not caused by a confounding event
Perform an event study to analyze if the company has abnormal returns
● Use varied event periods to identify the duration within which the company is affected by the event
● Analyze if the event's effect on share price varies between industries

Figure 2. Steps for analyzing information security events.

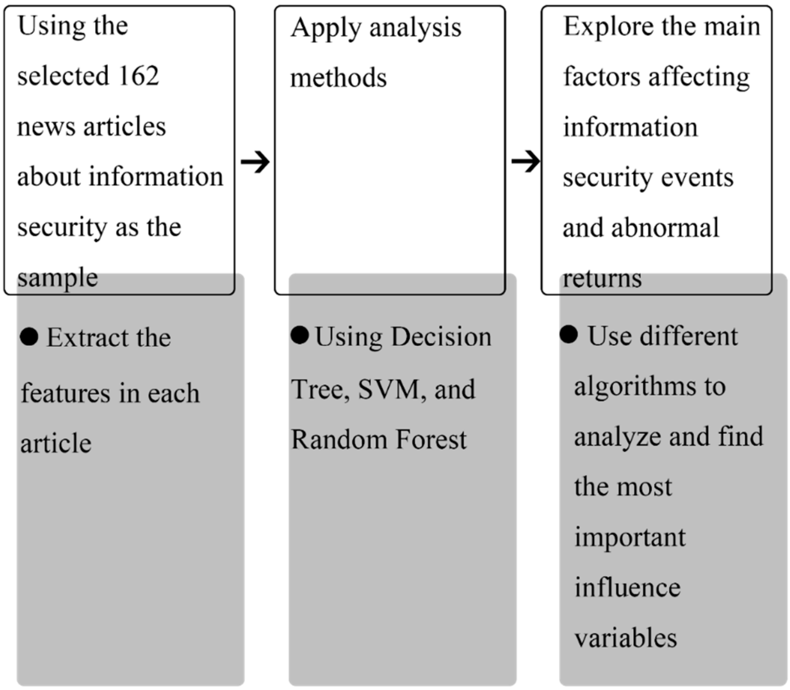

Using the selected 162 news articles about information security as the sample
● Extract the features in each article
Apply analysis methods
● Using Decision Tree, SVM, and Random Forest
Explore the main factors affecting information security events and abnormal returns
● Use different algorithms to analyze and find the most important influence variables

Figure 3. Steps for content analyzing information security events.

Table 3. The keywords used for searching news on information security events.

Computer Attack	Computer Worm	Identity Theft
Computer Breach	Cyber Fraud	Network Intrusion
Computer Break-in	Cyber-attack	Phishing
Computer Intrusion	Data Theft	Security Breach
Computer Security	Denial of Service	Security Controls
Computer Virus	Hacker	Security Incident

The selection of event samples followed five basic principles, as follows:

1. The firm involved in the event must belong to the S&P 500 index.
2. The event was reported in major newspapers in the U.S.
3. The firm has share trading records for 180 consecutive days before the event day.
4. Avoidance of the confounding effect, that is, the effect caused by another event that affects the estimation result, such as a corporate merger, release of financial statement, and turnover of high-ranking officers. Events affected by a confounding effect must be dropped to avoid confusion.
5. Event day is defined as the day on which the first article about the event is published.

3.3. The Calculation of Abnormal Returns

For an event study, there are numerous models for estimating the impact of events on share prices, among which the market model is most widely used. The market model was first applied to estimate the normal returns of firms and investigate how firms might be affected by general market factors. To apply this model, data within the estimation window are processed using the ordinary least squares (OLS) method to build a regression model for each stock [38]. The equation is as follows:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \tag{1}$$

where R_{it} is the rate of return on the stock price of firm i at event day t; $t = t_1, \ldots, t_2$; $i = 1, 2, \ldots, n$; α_i and β_i are parameters, ε_{it} is an error term, and R_{mt} denotes the market return in period t.

After the rate of return is estimated, event day i can be used to estimate the abnormal return with the following equation:

$$AR_{iE} = R_{iE} - (a_i + b_i R_{mE}) \tag{2}$$

where AR_{iE} represents the abnormal return of company i in event period E; a_i and b_i are parameters; R_{mE} denotes the market return at time period t; R_{iE} denotes the actual return of company i over the estimation period.

In the related literature, if the estimation model is built with daily return data, the estimation period is usually set between 100 and 300 days [5,6,39]. This study takes 180, a value approximately in the middle of this range, to be the length of the estimation period.

Next, with the abnormal returns obtained using the above-mentioned equations, the average abnormal return (AR_E) [39] is calculated as follows:

$$AR_E = \frac{1}{N}\sum_{i=1}^{N} AR_{iE} \tag{3}$$

where N denotes the number of firms.

Further, the available statistics are used to derive the cumulative abnormal return (CAR) from $\mathcal{T}_1$ to $\mathcal{T}_2$. The equation is as follows:

$$CAR(\mathcal{T}_1, \mathcal{T}_2) = \sum_{E=\mathcal{T}_1}^{\mathcal{T}_2} AR_E \tag{4}$$

3.4. Analysis of the Correlation between News Content and Abnormal Returns

The 162 news articles about information security were further analyzed to extract the features of each article, such as news source, presence of negative words in the headline, and number of negative words in the article. Based on the extracted features, this study applies decision tree, SVM, and random forest methods to explore the key factors affecting information security events and abnormal returns, as well as the importance of each variable. The negative words used in the analysis are adapted from Loughran and McDonald Sentiment Word Lists. The features used in the analysis are listed in Table 4.

Table 4. The features of news articles used in this study.

News Source	Headline Length (Words)	Total Word Count
Number of keywords in the article	Presence of negative words in the headline	Number of negative words in the article
Negative words/total words	Presence of any keyword in the headline	

4. Research Results

Most prior studies of security-related events use data mainly from a period between 1998 and 2012. However, due to the rapid advancement of information technologies, the attack types and security issues are much different in the present day, and so are market reactions to security events. Hence, this study collected news on information security events that occurred during 2009–2015 for an event study. This study first calculated and tested the abnormal returns of firms reported in the 162 news articles. The average abnormal return on T+1 is negative, indicating the reporting of information security events causes negative abnormal returns. The decision tree analysis shows that the news source and presence of negative words are critically important factors that affect abnormal returns. In further SVM and random forest analyses, the number of negative words, presence of negative words in the headline, and total word count, are also important variables that influence the effect of a news event on abnormal returns.

4.1. Data Analysis

Event Sample Selection and Classification

In the selection of events, the aforementioned keywords about information security were used to search for news articles published in major newspapers in the U.S. on Factiva. These news sources are listed in Table 5 [36].

Table 5. The news sources adopted in this study.

The New York Times	Washington Post	San Jose Mercury News
USA Today	Los Angeles Times	The Washington Post
New York Daily News	New York Post	Chicago Tribune
Chicago Sun-Times	The Denver Post	The Wall Street Journal
The Dallas Morning News	Newsday	The Orange County Register
Houston Chronicle		

Table 6 provides the statistics of the collected news articles. A total of 3846 news articles were found to contain one or several of these keywords. Excluding 761 repetitive articles, the preliminary sample consisted of 3085 news articles [36].

Table 6. The number of news articles after preliminary screening.

Keywords	Total Number of News Articles	Number of Usable News Articles	Number of Repetitive News Articles
Computer Attack	26	20	6
Computer Breach	16	15	1
Computer Break-in	0	0	0
Computer Intrusion	13	10	3
Computer Security	668	529	139
Computer Virus	63	50	13
Computer Worm	30	22	8
Cyber Fraud	1	1	0
Cyber-attack	157	142	15
Data Theft	134	110	24
Denial of Service	200	156	44
Hacker	964	786	178
Identity Theft	753	591	162
Network Intrusion	1	1	0
Phishing	326	256	70
Security Breach	427	348	79
Security Controls	46	33	13
Security Incident	21	15	6
Total	3846	3085	761

Afterwards, each news article was evaluated to determine if it belongs to an information security event. As the subjects reported in the news are S&P firms, it is necessary to check the presence of any confounding event in each reported event. The screening procedure is as follows:

1. Search on Factiva for announcements of any major news about the company within a certain period of time before and after the event day. Events such as corporate merger, release of financial statements or turnover of high-ranking executives may cause an effect on a company's share prices. The event should be dropped if any confounding event is detected.
2. Use Google Search to make sure again that the company has no other incident around the event day.

4.2. Analysis Results

Calculating Abnormal Returns

This study first calculated and tested the abnormal returns of firms reported in the 162 news articles. With the event window set as [−10,10] and the estimation period set as 180 days, the results as shown in Table 7 [36]. As shown in this table, the average abnormal return on $T+1$ is −0.00214, indicating the reporting of information security events causes negative abnormal returns. The values for $T+6$, $T+8$, and $T+9$ are also slightly significant. Because several days have passed after the event day, and the stock price has become much stable on the second day after the event, the abnormal returns on these days might have been affected by other news events or a potential confounding event. Hence, the statistics for these days should not be considered.

Table 7. Average abnormal returns, t-test, and standardized residual test results.

Event Day	Number of News Articles	Average Abnormal Return (%)	Standard Deviation
−10	162	0.000953	0.017391
−9	162	−0.00057	0.012463
−8	162	−0.00053	0.01217
−7	162	0.000293	0.0127
−6	162	−0.00051	0.016119
−5	162	−0.00167	0.01615
−4	162	0.001379	0.013438
−3	162	−0.00012	0.012967
−2	162	−0.0002	0.011417
−1	162	−0.00044	0.014416
0	162	0.0002	0.014837
1	162	−0.00214	0.010874
2	162	0.00174	0.014141
3	162	−0.00068	0.01466
4	162	−0.00029	0.014945
5	162	−0.0017	0.013769
6	162	−0.00202	0.017384
7	162	−0.00114	0.012772
8	162	0.001956	0.015479
9	162	−0.00216	0.013369
10	162	−0.00062	0.013967

4.3. Analyzing News Content and Abnormal Returns

Before applying algorithms to analyze the news sample, the news sources are coded first. The code table for news sources is presented in Table 8.

Table 8. The code table for news sources.

N01	USA Today	N02	The Orange County Register
N03	The Wall Street Journal	N04	San Jose Mercury News
N05	Houston Chronicle	N06	The New York Times
N07	Newsday	N08	New York Post
N09	Los Angeles Times	N10	The Washington Post
N11	Chicago Tribune	N12	Washington Post
N13	New York Daily News		

(1) Using Decision Tree to Analyze the Features in News Content

By applying the decision tree C5.0 algorithm, the classification result for Node 0 is shown in Table 9. In this node, 84 news events mentioned abnormal returns, and 78 did not. The importance of each variable is as shown in Figure 4. Among these variables, the number of keywords in the article is the most important, followed by presence of negative words in the headline and the number of negative words in the article. It can be inferred that presence of negative words in information security news articles is related to abnormal returns of firms.

Table 9. The classification result for Node 0 in the decision tree.

Node 0		
Category	**%**	**n**
<0%	51.852	84
>=0%	48.148	78
Total	100.000	162

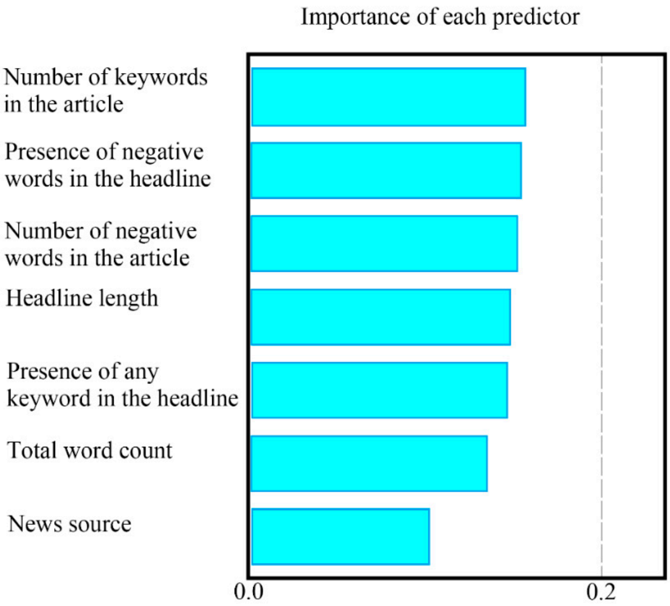
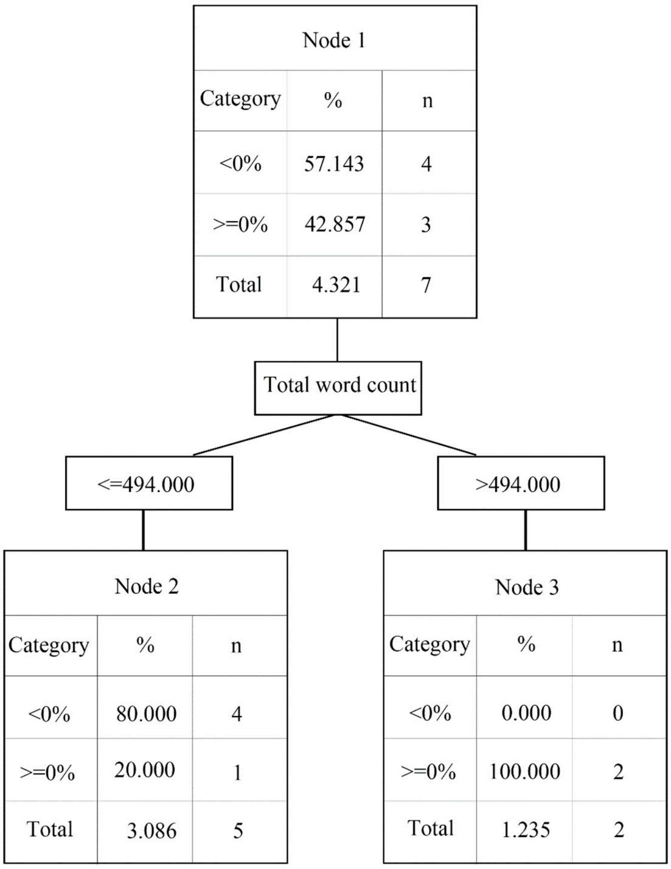


Figure 4. The importance of each predictor in the decision tree analysis.

Figure 5 shows the classification results at Node 1, Node 2, and Node 3 for news source N01 (USA Today). In this news source, total word count is a relatively more important variable. At Node 2, articles with a word count <=494 may cause abnormal returns; at Node 3, those with a word count >494 are not associated with abnormal returns. Below are the classification results for news source N01 (USA Today).



Node 1

Category	%	n
<0%	57.143	4
>=0%	42.857	3
Total	4.321	7

Node 2

Category	%	n
<0%	80.000	4
>=0%	20.000	1
Total	3.086	5

Node 3

Category	%	n
<0%	0.000	0
>=0%	100.000	2
Total	1.235	2

Figure 5. The classification results for N01 (USA Today).

At Node 4, six news sources including N02 (The Orange County Register), N05 (Houston Chronicle), N07 (Newsday), N11 (Chicago Tribune), N12 (Washington Post), N13 (New York Daily News) were found to have no significant association with abnormal returns. The decision tree classification results for N02, N05, N07, N11, N12, and N13 are presented in Table 10.

Regarding the classification results at nodes from Node 5 to Node 11, news source N03 (The Wall Street Journal) has 58 news articles, constituting a relatively large proportion compared to other news sources. At Node 6, word count <=1139 has an effect on abnormal returns. At Node 11, word count>1139 is not associated with abnormal returns. Below Node 6 (<=1139) are two branches, namely Node 7 and Node 8. At Node 7, the number of negative words <=14 may lead to abnormal returns. At Node 8, the number of negative words >14 is not associated with abnormal returns. The classification results are detailed in Figures 6 and 7. Figure 7 shows the classification results at Node 9 and Node 10 that are split from the upper-level node.

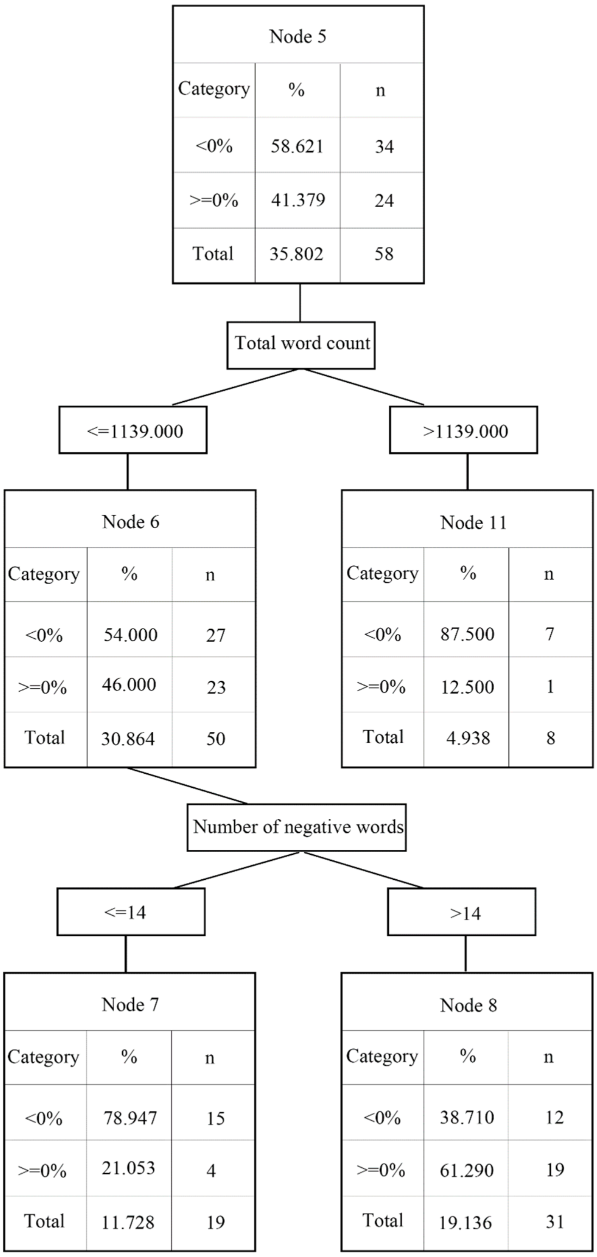


Figure 6. The classification results for N03 (The Wall Street Journal) at Nodes 5, 6, 7, 8, and 11.

Table 10. The classification results for N02, N05, N07, N11, N12, and N13.

Node 4		
Category	**%**	**n**
<0%	16.667	3
>=0%	83.333	15
Total	11.111	18

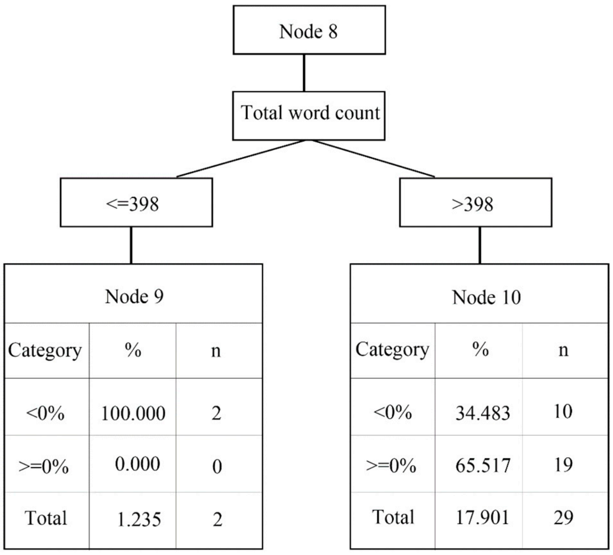


Figure 7. The classification results for N03 (The Wall Street Journal) at Node 9 and Node 10.

For news source N04 (San Jose Mercury News), the result at Node 12 is as shown in Table 11, which indicates no variable has a significant effect on abnormal returns.

Table 11. The classification results for N04 (San Jose Mercury News) at Node 12.

Node 12		
Category	**%**	**n**
<0%	60.000	3
>=0%	40.000	2
Total	3.086	5

From Node 13 to Node 15, news source N06 (The New York Times) has 38 news articles at Node 13. At Node 14, headline length (words) <=6 leads to abnormal returns. At Node 15, headline length (words) >6 is not as associated with abnormal returns. The results are as shown in Figure 8.

From Node 16 to Node 18, news source N08 (New York Post) has four articles. The results at Node 17 and Node 18 indicate that total word count <=496 is related to abnormal returns. The details are provided in Figure 9.

Nodes 19–21 present the classification results for N09 (Los Angeles Times) which makes up 18 news articles. At Node 17, the number of negative words <=29 may cause abnormal returns. At Node 18, the number of negative words >29 is not associated with abnormal returns. The details are provided in Figure 10.

Nodes 22–26 show the classification results for N10 (The Washington Post) which has 14 news articles in the sample. The result at Node 23 shows that presence of negative words in the headline (=Yes) is not associated with abnormal returns. This suggests that headlines that contain negative words may not cause significant abnormal returns. Below Node 23 are two branches, namely Node 24 and Node 25, that further test the effect of headline length. At Node 24, headline length (words)

<=17 is not associated with abnormal returns. At Node 25, headline length (words) >17 may cause abnormal returns. The details are provided in Figure 11.

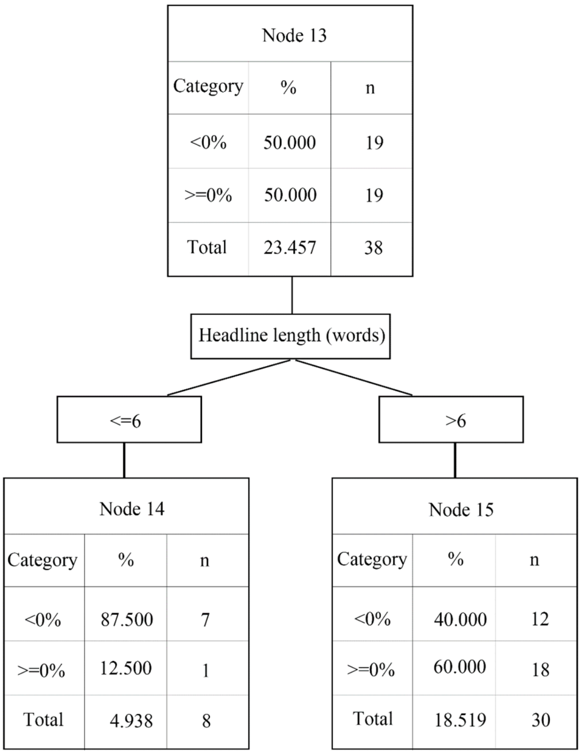


Figure 8. The classification results for N06 (The New York Times) at Nodes 13–15.

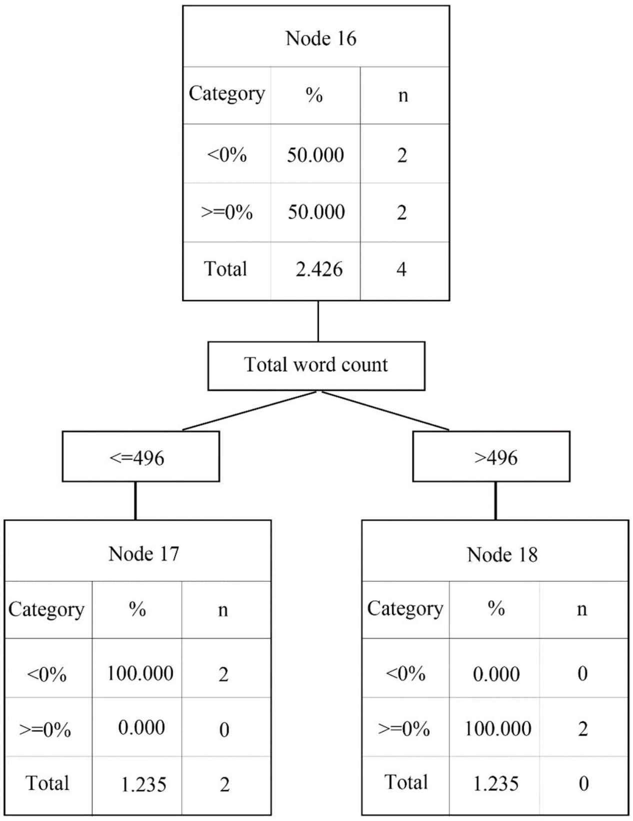


Figure 9. The classification results for N08 (New York Post) at Nodes 16–18.

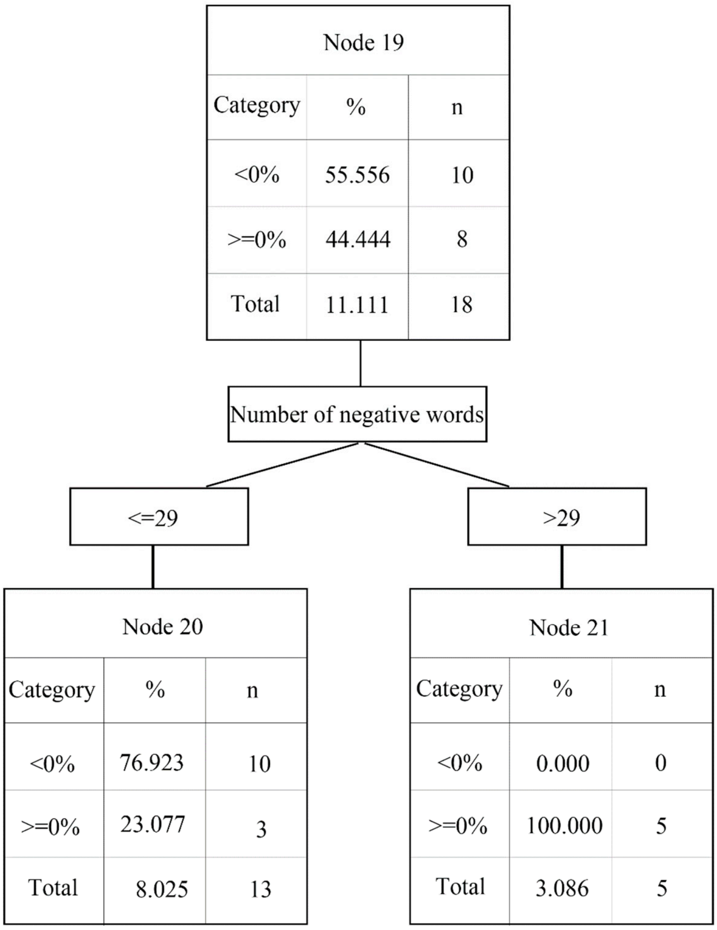


Figure 10. The classification results for N09 (Los Angeles Times) at Nodes 19–21.

Through the analysis of the nodes to understand the different news sources, the affected factors will also be different; for example, USA Today's main impact factor is total word count.

(2) Using SVM to Analyze the Features of News Context

The SVM analysis evaluates the importance of each variable as shown in Figure 12. This figure shows that the news source plays an important role in determining the effect of an information security event on abnormal returns. In other words, some media sources are more influential on the public than others. The presence of negative words in the headline is another variable with a notable influence. Headlines that contain negative words cause negative sentiments for readers, which in turn lead to abnormal returns. The other variables exert similar degrees of influence on abnormal returns.

(3) Random Forest Analysis of Features of News Content

The random forest decision-making rules are listed in Table 12.

Figure 13 shows the importance of each predictor obtained from a random forest analysis. As can be seen in this figure, total word count, headline length, news source, number of negative words in the article, and negative words/total words are the five variables that exert a greater influence on abnormal returns. Total word count and headline length have always been pivotal to a reader's decision to read a news article. In addition, mainstream news media are the major sources of news for most people. When reading a news article containing more negative words in the article, readers definitely have more concerns about the operation of the reported firm. The other factors have similar effects on abnormal returns.

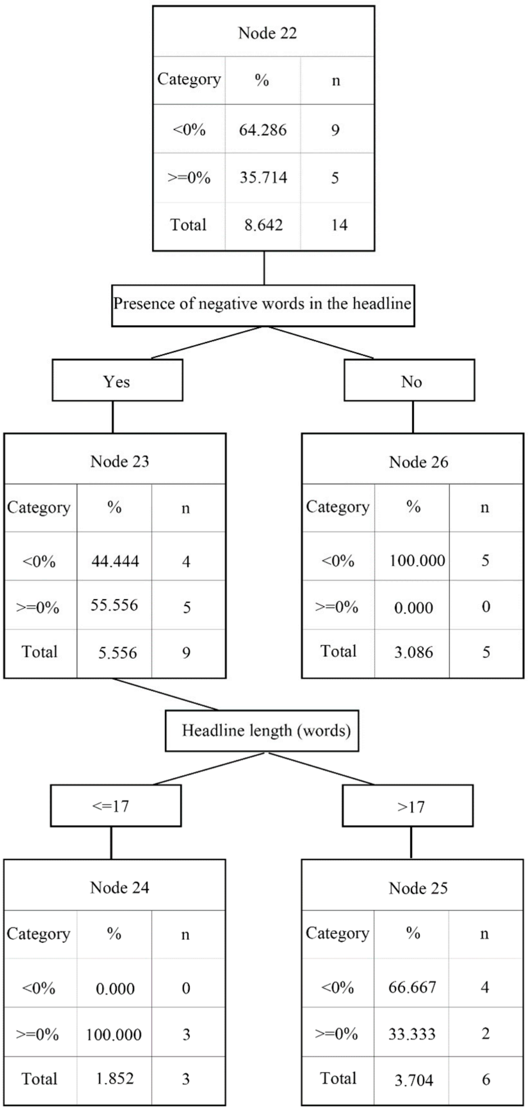


Figure 11. The classification results for N10 (The Washington Post) at Nodes 22–26.

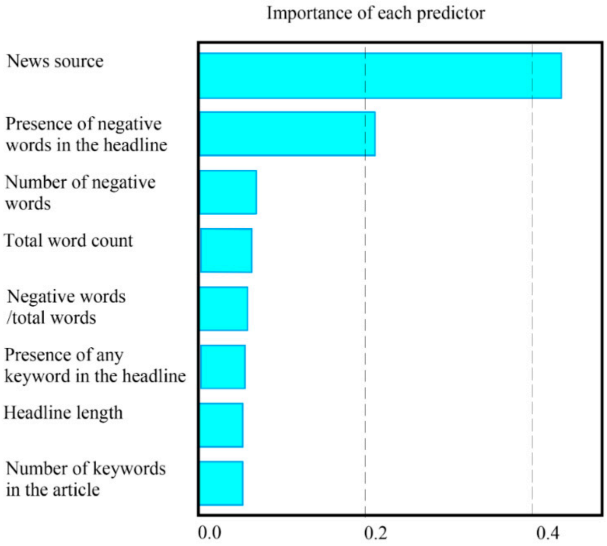


Figure 12. Support vector machine (SVM) analysis of the importance of each variable for abnormal returns.

Table 12. The decision rules in the random forest.

"Abnormal Return" is the Highest Decision Rule for the Following Items				
Decision Rule	**The Most Common Type**	**Rule Precision**	**Forest Precision**	**Interest Index**
(Number of negative words in the article <= 38.0) and (News source = {N01,N02,N03,N04,N05,N06,N08,N09,N10,N11,N12,N13}) and (Negative words/total word count > 3.633) and (Headline length (words) <= 13.0) and (Presence of negative words in the headline={No})	<0%	1.000	1.000	1.000
(Total word count > 1215.0) and (Headline length (words) <= 23.0) and (News source = {N03,N04,N08}) and (News source = {N01,N03,N04,N08,N10})	<0%	1.000	1.000	1.000
(Number of negative words in the article <= 38.0) and (Negative words/total word count <= 3.633) and (Headline length (words) <= 9.0) and (Presence of negative words in the headline ={No}) and (Headline length (words) > 6.0)	<0%	1.000	1.000	1.000
(Headline length (words) <= 6.0) and (Number of negative words in the article <= 38.0) and (Number of negative words in the article > 11.0) and (News source = {N04,N06}) and (News source = {N01,N04,N06,N08,N09,N10})	<0%	1.000	1.000	1.000
(Number of negative words in the article > 38.0) and (Number of negative words in the article > 11.0) and (News source = {N04,N06}) and (News source = {N01,N04,N06,N08,N09,N10})	>=0%	1.000	1.000	1.000

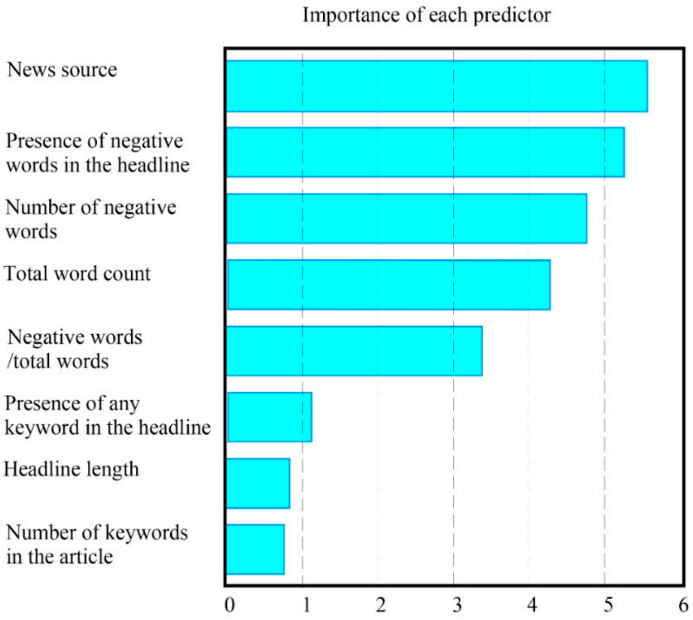


Figure 13. Random forest analysis of the effect of each variable on abnormal returns.

5. Conclusions

In this study, the information security events published in major newspapers in the U.S. between 1 January 2009 and 31 December 2015 are analyzed. The conclusions are summarized as follows:

(1) The empirical evidence indicates that news coverage of corporate information security events diffuses negative messages among investors, which would in turn cause fluctuation in the firm's share prices and generation of negative returns.

(2) The decision tree analysis shows that the news source and negative words are critically important factors that affect abnormal returns.

(3) In further SVM and random forest analyses, other factors are examined, including the number of negative words, presence of negative words in the headline, and total word count. These are found to also be important variables that influence the effect of a news event on abnormal returns.

Through the analyses and statistical testing of the event study method, information security news has significant influence on firms and may lead to negative returns.

The results can contribute to higher corporate awareness of the importance of information security tasks, from regular education training of employees to strengthening of corporate information systems. According to the findings, firms are advised to find and develop preventive methods and solutions to achieve better and more comprehensive protection of information security. Firms, especially those in industries characterized by relatively higher market sensitivity, need to ensure information security so as to avoid losses resulting from a security breach.

Finally, despite efforts to collect a more representative sample, news about information security can be diffused not only via news media but also via other channels, such as social media. However, a more appropriate model for collecting and analyzing samples from multiple media is still absent. Whether investors are influenced by news from social media to make a different investment decision cannot be predicted. The effect of this news diffusion channel must be considered and analyzed in future research.

Funding: This research received no external funding.

Acknowledgments: I would like to thank the editor and two anonymous reviewers for their constructive comments, which helped me to improve the manuscript.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Mohr, J.J. The management and control of information in high-technology firms. *J. High Technol. Manag. Res.* **1996**, *7*, 245–268. [CrossRef]
2. Arya, B.; Zhang, G. Institutional Reforms and Investor Reactions to CSR Announcements: Evidence from an Emerging Economy. *J. Manag. Stud.* **2009**, *46*, 1089–1112. [CrossRef]
3. Carow, K.; Heron, R.; Saxton, T. Do early birds get the returns? An empirical investigation of early-mover advantages in acquisitions. *Strateg. Manag. J.* **2004**, *25*, 563–585. [CrossRef]
4. Chen, Y.; Ganesan, S.; Liu, Y. Does a Firm's Product-Recall Strategy Affect Its Financial Value? An Examination of Strategic Alternatives during Product-Harm Crises. *J. Mark.* **2009**, *73*, 214–226. [CrossRef]
5. Konchitchki, Y.; O'Leary, D.E. Event study methodologies in information systems research. *Int. J. Account. Inf. Syst.* **2011**, *12*, 99–115. [CrossRef]
6. Wang, T.; Ulmer, J.R.; Kannan, K. The Textual Contents of Media Reports of Information Security Breaches and Profitable Short-Term Investment Opportunities. *J. Organ. Comput. Electron. Commer.* **2013**, *23*, 200–223. [CrossRef]
7. Huang, C.K.; Wang, T.T.; Tsai, Y.T. Market reactions to big data implementation announcements. In Proceedings of the Pacific Asia Conference on Information Systems, Chiayi, Taiwan, 27 June–1 July 2016.
8. Dos Santos, B.L.; Peffers, K.; Mauer, D.C. The Impact of Information Technology Investment Announcements on the Market Value of the Firm. *Inf. Syst. Res.* **1993**, *4*, 1–23. [CrossRef]
9. Son, I.; Lee, D.; Lee, J.N.; Chang, Y.B. Market perception on cloud computing initiatives in organizations: An extended resource-based view. *Inf. Manag.* **2014**, *51*, 653–669. [CrossRef]
10. Rubin, E.; Rubin, A. The impact of Business Intelligence systems on stock return volatility. *Inf. Manag.* **2013**, *50*, 67–75. [CrossRef]
11. Songur, H.; Heavilin, J.E. Abnormal research and development investments and stock returns. *North Am. J. Econ. Finance* **2017**, *42*, 237–249. [CrossRef]
12. Song, X.; Tippett, M.; Vivian, A. Assessing abnormal returns: The case of Chinese M & A acquiring firms. *Res. Int. Bus. Finance* **2017**, *42*, 191–207.

13. Modi, S.B.; Wiles, M.A.; Mishra, S. Shareholder value implications of service failures in triads: The case of customer information security breaches. *J. Oper. Manag.* **2015**, *35*, 21–39. [CrossRef]
14. Leitch, D.; Sherif, M. Twitter mood, CEO succession announcements and stock returns. *J. Comput. Sci.* **2017**, *21*, 1–10. [CrossRef]
15. Siponen, M.T. Six Design Theories for IS Security Policies and Guidelines. *J. AIS* **2006**, *7*, 19. [CrossRef]
16. Straub Jr, D.W. Effective IS security: An. empirical study. *Inf. Syst. Res.* **1990**, *1*, 255–276. [CrossRef]
17. Gal-Or, E.; Ghose, A. The Economic Incentives for Sharing Security Information. *Inf. Syst. Res.* **2005**, *16*, 186–208. [CrossRef]
18. Schechter, S.E.; Smith, M.D. How Much Security Is Enough to Stop a Thief? In *Financial Cryptography, Proceeding of the 7th International Conference, Guadeloupe, French, 27–30 January 2003*; Wright, R.N., Ed.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 122–137.
19. Kwon, J.; Ulmer, J.R.; Wang, T. The Association between Top Management Involvement and Compensation and Information Security Breaches. *J. Inf. Syst.* **2013**, *27*, 219–236. [CrossRef]
20. Gordon, L.A.; Loeb, M.P.; Sohail, T. Market value of voluntary disclosures concerning information security. *MIS Q.* **2010**, *34*, 567–594. [CrossRef]
21. Kositanurit, B.; Osei-Bryson, K.M.; Ngwenyama, O. Re-examining information systems user performance: Using data mining to identify properties of IS that lead to highest levels of user performance. *Expert Syst. Appl.* **2011**, *38*, 7041–7050. [CrossRef]
22. Feldman, R.; Sanger, J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*; Cambridge University Press: Cambridge, UK, 2006.
23. Han, J.; Altman, R.B.; Kumar, V.; Mannila, H.; Pregibon, D. Emerging scientific applications in data mining. *Commun. ACM* **2002**, *45*, 54–58. [CrossRef]
24. Cecchini, M.; Aytug, H.; Koehler, G.J.; Pathak, P. Detecting Management Fraud in Public Companies. *Manag. Sci.* **2010**, *56*, 1146–1160. [CrossRef]
25. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
26. Pal, R.; Kupka, K.; Aneja, A.P.; Militky, J. Business health characterization: A hybrid regression and support vector machine analysis. *Expert Syst. Appl.* **2016**, *49*, 48–59. [CrossRef]
27. Niklis, D.; Doumpos, M.; Zopounidis, C. Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines. *Appl. Math. Comput.* **2014**, *234*, 69–81. [CrossRef]
28. Al-Yaseen, W.L.; Othman, Z.A.; Nazri, M.Z.A. Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. *Expert Syst. Appl.* **2017**, *67*, 296–303. [CrossRef]
29. Khalifi, H.; Elqadi, A.; Ghanou, Y. Support Vector Machines for a new Hybrid Information Retrieval System. *Procedia Comput. Sci.* **2018**, *127*, 139–145. [CrossRef]
30. Bellazzi, R.; Zupan, B. Predictive data mining in clinical medicine: Current issues and guidelines. *Int. J. Med. Inf.* **2008**, *77*, 81–97. [CrossRef]
31. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
32. Xie, Y.; Li, X.; Ngai, E.W.T.; Ying, W. Customer churn prediction using improved balanced random forests. *Expert Syst. Appl.* **2009**, *36*, 5445–5449. [CrossRef]
33. Whitrow, C.; Hand, D.J.; Juszczak, P.; Weston, D.; Adams, N.M. Transaction aggregation as a strategy for credit card fraud detection. *Data Min. Knowl. Discov.* **2009**, *18*, 30–55. [CrossRef]
34. Kouzani, A.Z. Faceparts for Recognition. In Proceedings of the TENCON 2006—2006 IEEE Region 10 Conference, Hong Kong, China, 14–17 November 2006; pp. 1–4.
35. Gupta, R.; Pierdzioch, C.; Vivian, A.J.; Wohar, M.E. The predictive value of inequality measures for stock returns: An analysis of long-span UK data using quantile random forests. *Financ. Res. Lett.* **2018**. [CrossRef]
36. Hung, C.C.; Huang, C.K.; Ku, C.Y. Research on Abnormal Return of Enterprise Stock Price for the Information Security News. *J. Inf. Manag.* **2018**, *25*, 283–306.
37. Campbell, K.; Gordon, L.A.; Loeb, M.P.; Zhou, L. The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *J. Comput. Secur.* **2003**, *11*, 431–448. [CrossRef]

38. Sharpe, W.F. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *J. Finance* **1964**, *19*, 425–442.
39. McWilliams, A.; Siegel, D. Event Studies in Management Research: Theoretical and Empirical Issues. *Acad. Manag. J.* **1997**, *40*, 626–657.