Association for Information Systems
**AIS Electronic Library (AISeL)**

PACIS 2017 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

Summer 7-19-2017

# Integration of Social Media News Mining and Text Mining Techniques to Determine a Corporate's Competitive Edge

Te-Ming Chang
*National Sun Yat-sen University*, temin@cm.nsysu.edu.tw

Guo Hsin Hu
*National Sun Yat-Sen University*, guohsin@mail.mirdc.org.tw

Ming-Fu Hsu
*Chinese Culture University*, hsumf0222@gmail.com

Keng-Pei Lin
*National Sun Yat-Sen University*, kplin@mis.nsysu.edu.tw

# Integration of Social Media News Mining and Text Mining Techniques to Determine a Corporate's Competitive Edge

**Te-Min Chang**
Department of Information Management, National Sun Yat-sen University, 70 Lienhai Rd., Kaohsiung 80424, Taiwan, R.O.C.
temin@cm.nsysu.edu.tw

**Guo-Hsin Hu**
Department of Information Management, National Sun Yat-sen University, Metal Industries Research & Development Centre, 70 Lienhai Rd., Kaohsiung 80424, Taiwan, R.O.C.
guohsin@mail.mirdc.org.tw

**Ming-Fu Hsu***
English Program of Global Business, Chinese Culture University, 55, Hwa-Kang Rd., Yang-Ming-Shan, Taipei , 11114, Taiwan, R. O. C.
*(Corresponding author: M.-F. Hsu)
hsumf0222@gmail.com

**Keng-Pei Lin**
Department of Information Management, National Sun Yat-sen University, 70 Lienhai Rd., Kaohsiung 80424, Taiwan, R.O.C.
kplin@mis.nsysu.edu.tw

## Abstract

*Market globalization have triggered much more severe challenges for corporates than ever before. Thus, how to survive in this highly fluctuating economic atmosphere is an attractive topic for corporate managers, especially when an economy goes into a severe recession. One of the most consensus conclusions is to highly integrate a corporate's supply chain network, as it can facilitate knowledge circulation, reduce transportation cost, increase market share, and sustain customer loyalty. However, a corporate's supply chain relations are unapparent and opaque. To solve such an obstacle, this study integrates text mining (TM) and social network analysis (SNA) techniques to exploit the latent relation among corporates from social media news. Sequentially, this study examines its impact on corporate operating performance forecasting. The empirical result shows that the proposed mechanism is a promising alternative for performance forecasting. Public authorities and decision makers can thus consider the potential implications when forming a future policy.*

**Keywords:** Text mining, social network analysis, decision making, forecasting

## Introduction

The recent financial crisis and its subsequent turmoil that erupted in 2007 have once again highlighted the essence of precise forecasting, especially in today's highly fluctuating economic atmosphere. A financial crisis not only can cause a large amount of economic and social losses for all stakeholders, but can also destroy corporates themselves. Consequently, due to its considerable economic influence on decision making, the establishment of a financial pre-warning model has become an attractive topic over the decades. If the forecasting model is trustful, then higher-level managers can initiate some strategies or treatments to prevent the financial crisis from bursting out, and investors can grasp the opportunity to modify their investment portfolios so as to maximize their personnel wealth under any anticipated risk exposure.

The pioneer work by Altman (1968) used multivariate discriminant analysis (MDA) to construct a discriminant function, called the Z-score, for financial crisis prediction. If the Z-score is greater than 2.99, then a corporate is classified into the non-financial crisis group; if the Z score is less than 1.81, then a corporate is classified into the financial crisis group. Ohlson (1980) employed logistic regression (LR) to construct the forecasting model, with the result showing a satisfactory forecasting performance. Wilson and Sharda (1994) implemented neural network (NN) to deal with the financial crisis forecasting task, finding that NN outperforms DA in terms of forecasting accuracy. Support vector machine (SVM) was employed by Kim and Sohn (2010) to construct a financial crisis forecasting model. The result showed that SVM outperforms the other two classifiers (i.e., NN and LR). However, when compared with well-examined studies such as financial crisis prediction, and credit risk prediction, research on corporate operating performance forecasting is quite rare. Kamei (1997) stated that 99% of financial crises are due to bad corporate operating performances. In other words, events such as defaults on promissory notes and corporate insolvency do not just take place suddenly. Instead, there are notable root causes that precede a corporate financial crisis, and it is the inability to deal with such events properly at an early stage that results in the demise of these corporates (Shirata and Sakagami, 2008). In brief, a corporate operating badly happens in an irreversible prior stage before a financial crisis erupts.

According to previous work done by Ghalayini and Noble (1996), a corporate operating performance measure can be briefly divided into two stages: in the first stage, which went on until the 1980s, the performance measures were restricted to financial ratios supplied by the management accounting system; the second stage, which began in the late 1980s and is still proceeding, brought about many modifications within the performance measure, such as customer satisfaction, economic value added, internal operations performance, and intangible assets, and attention in this domain has increased considerably. One of the most well-known methods for this purpose was balanced scorecard (BSC) introduced by Kaplan and Norton (1996). BSC provides an intuitive and comprehensive performance measure method that concentrates on financial and non-financial perspectives, long-term and short-term strategies, and internal and external business assessments, as well as it can help a corporate to navigate the business environment to reach feature competitive success. Based on the aforementioned merits, BSC has become an essential method for performance measurement. Although BSC has numerous merits, it also comes with a critical challenge: it is impossible to reach the final performance rank. To solve this problem, this present study implements an non-parametric algorithm, called data envelopment analysis (DEA), which can produce the efficiency frontier based on the concept of the Pareto optimum (Guo and Zhu, 2017; Cook et al., 2017). By combining BSC and DEA, we can acquire the final performance rank that consists of overarching viewpoints of corporate operations, which are then fed into a kernel-based technique to construct the model for corporate operating performance forecasting.

Support vector machine (SVM), a kernel-based technique, has caught considerable attention by researchers due to its superior generalization ability, higher operating efficiency, and better excessive matching capability (Sachdeva et al., 2016; Yin and Yin, 2016). However, given $k$ training samples, calculating the corresponding kernel matrix by utilizing a quadratic programming (QP) solver takes up $O$ $(k^3)$ training time and at least $O$ $(k^2)$ space (Collobert et al., 2002), thus heavily restricting practical application to a large-scale sample data. To handle the aforementioned obstacle, Tsang et al. (2005) introduced core vector machine (CVM), which has asymptotic linear time complexity with $k$ , and its space complexity is independent of $k$ . It converts the QP problem encountered by SVM into a minimum enclosing ball (MEB) problem and sequentially implements an iterative $(1+\varepsilon)$ -approximation method to deal with the MEB problem. By doing so, it can be used to deal with a forecasting task having large-scale sample data and complicated non-linearity (Tsang et al., 2005; Gan et al., 2013). To our knowledge, no current study has executed CVM to construct the model for corporate operating performance forecasting. To fill this gap, this study takes CVM to construct the pre-warning model.

The current status of the global economic atmosphere is highly fluctuating and full of uncertainty. How to facilitate and sustain a corporate's competitive edge under this business environment is an essential topic for decision makers. Although coming up with consensus conclusions is very complicated, it is widely accepted that corporates' business relationship networks (such as supply chain relationships, cooperation relationships, and alliance relationships) are one of the prescribed avenues to strengthen their competitive capability and upgrade their industrial level (Shafiee et al., 2014; Chang et al., 2015). The rationale is that involvement in its business relationship network can assist a corporate in responding to customers' requirements and market changes, increase customer loyalty and value, and enlarge market share and profit margins, by making use of this network for the transfer of valuable knowledge, information, and opportunities (Acquaah, 2011; Kim, 2014; Naudé et al., 2014; Kim et al., 2016). Dyer and Nobeoka (2000) also stated that a corporate having a higher level of involvement in its business relationship network can considerably reduce transaction costs. However, the related works on real-life corporate business relationship network identification is quite rare, because the nature of a corporate's business relationship is unapparent and opaque (Ma et al., 2011). Thus, decision makers have to spend a considerable amount of time and money to identify their corporate's business relationships. Manually determining a corporate's business relationship network is a very time-consuming and impractical work. To handle the aforementioned task, this study implements a text mining (TM) technique to extract the knowledge from seemingly noisy social media news and uses structural features grounded on a social network analysis (SNA) technique to determine a corporate's competitive edge (i.e., the corporate's position embedded into its business relationship network). This study also not only examines the influence of a corporate's competitive edge on performance assessment, but also uses it to construct the model for corporate operating performance forecasting. Our main contributions of this study can be summarized as follows.

- We introduce an effective and efficient model for corporate operating performance forecasting, as poor corporate performance has been widely deemed as the main cause of a financial crisis.
- We implement a TM technique to construct the corporate business relationship network by using seemly noisy social media news and then extract the inherent knowledge from the network by a SNA technique.
- The network can be used by public sectors to identify which corporates present an outstanding competitive priority without involving a large cost.
- Validated by real-life samples, the introduced model is a promising alternative for corporate operating performance forecasting.

The rest of this paper is organized as follows. Section 2 reviews the extant literature. Section 3 provides as illustration of data and the introduced model. Section 4 summarizes the experimental results and statistical tests. Finally, Section 5 draws the conclusions.

## Literature review

A financial pre-warning model is a monitoring and reporting mechanism that provides alerts for the possibility of financial troubles, obstacles, and risks before they affect the corporate's financial situation (Li et al., 2014; Geng et al., 2015; Chang and Hsu, 2016; Fallahpour et al., 2017). It also provides an opportunity for high-level managers to mitigate or avoid potential fatal shocks in this highly fluctuating economic atmosphere. Most related works concentrate on financial distress/bankruptcy prediction, e.g., with the univariate approach (Beaver, 1966), multiple regression (Meyer and Pifer, 1970), logistic regression (Dimitras et al., 1996), and stepwise regression (Laitinen and Laitinen, 2000), and focus exclusively on monetary-based information, such as financial statements and income statements. However, compared with well-studied works, such as financial distress prediction and financial bankruptcy prediction, the works on corporate performance forecasting is quite rare. Furthermore, previous works on performance forecasting mainly concentrate on monetary-based information, such as return on assets (ROA) and return on equities (ROE), to determine a corporate's operating performance, which has been criticized as the root cause of many problems (Hafeez et al., 2002). Merely utilizing monetary-based information cannot depict the whole aspects of a corporate's operating situation. For example, as high-level managers place much more emphasis on short-term financial performance metrics, they have a tendency to trade away actions like production improvement, human resource development, customer royalty retention, and research and development expenditure that can bring in long-term advantages for current profitability, thus limiting the number of investments with future growth potential and opportunities (Banker et al., 2004).

In an attempt to deal with the aforementioned challenge by supplementing monetary-based measures with additional information (i.e., non-monetary-based measures) that can assist in assessing the long-term development of corporates, Kaplan and Norton (1992) provided the balanced scorecards (BSC), which is a performance evaluation mechanism that provides overarching viewpoints on corporate operating performance assessment by a set of measures that contain both financial and non-financial metrics (Kaplan and Norton, 1996). The fundamental idea of BSC is to maintain a balance between short- and long-term objectives, between lagging and leading indicators, between internal and external performance aspects, and between financial and non-financial measures (Kaplan and Norton, 1996). It also provides an understanding of high-level managers' strategies and translates them into a series of strategy objectives with operational measures at a lower level (Shen et al., 2016).

Due to numerous merits of BSC, it turns out to be the most well-known approach for performance measurement and has been widely utilized in numerous research domains with satisfactory outcomes, such as supply chain application (Shafiee et al., 2014), outsourcing decision making (Tjader et al., 2014), banking institutions (Wu, 2016), and sustainable technology selection (Xia et al., 2017).

Although multiple perspectives and measures embedded into BSC may reliably and appropriately represent the multifaceted nature of the corporate operation, they also cause some difficulties: (1) the need to account for mutually inconsistent outcomes, (2) the possible need to assign non-equal priorities to perspectives or measures, and (3) the need to design an aggregated metric that would somehow summarize the whole story of success (Bentes et al., 2012). One of the approaches that can address the complicated issues of a balanced system of performance evaluation is data envelopment analysis (DEA), which was introduced by Charnes et al. (1978). It is a mathematical linear programming approach that can be utilized to assess relative efficiency among a homogeneous set of decision-making units (DMUs) and gauge productivity without requiring any pre-determined production function. It can also be used to aggregate multiple features affecting corporate operating performance into a synthesized and composited performance rank. Thus, it has been considerably adopted to deal with the performance assessment task in both public and private organizations, including the hotel industry (Hwang and Chang, 2003), banking industry (Banker et al., 2010), construction industry (Xue et al., 2015), and universities (Sagarra et al., 2017). In order to provide an overarching description on corporate operations, this study conducts the integration of BSC and DEA.

With the great development of international trade and market globalization, local corporates nowadays are facing much more severe competition from multinationals and foreign corporates than ever before. In comparison with foreign corporates, local corporates invest relatively less in research and development (R&D) and normally encounter a large proportion of uncertainty in light of limited economic resources and information gaps in their technological know-how for innovation (Kim, 2014). This situation constricts the existence of corporates and even worse triggers bankruptcy. How to maintain and enhance a corporate's competitive privilege is an attractive research issue. One of the most widely accepted consensus conclusions is that a corporate highly involved in its supply chain network can incur a range of benefits, including reduced transportation costs, increased market share and profit margins, sustainable customer relationships, the elimination of market uncertainty, and innovation capability, by drawing on networks for missing resources (Ferguson, 2000; Qiao et al., 2014). Thus, the work on supply chain management has become an interesting topic to be discussed, such as green supply chain management (Sharma et al., 2017), supply chain coordination with information sharing (Zhou et al., 2017), and supply chain management with different inventory policies (Ponte et al., 2017). Unfortunately, as the nature of a supply chain relationship is unapparent, the existing works on supply chain relationship determination are very rare. To solve this challenge, this study adopts the TM technique and SNA technique. The TM technique is used to extract hidden information (i.e., supply chain relationship) from social media news and then the SNA technique is implemented to determine a corporate's competitive edge from the supply chain network. Finally, the performance rank decided by BSC+DEA and a corporate's competitive edge determined by TM+SNA are injected into a kernel-based technique to construct the model for corporate operating performance forecasting. If the forecasting model is trustful and reliable, then decision makers can initiate some strategy to avoid financial troubles from bursting out, and investors can modify their investment portfolio to reduce the anticipated risk level.

## Methodologies

### *Data envelopment analysis: DEA*

DEA was first introduced by Charnes et al. (1978) and grounded on the seminal work of Farrell (1957) for the assessment of productive efficiency. The basic concept in this method is to maximize the ratio of weighted output variables to weighted input variables for a group of units (called decision-making units: DMUs). It has two different versions: the DEA-CCR (Charnes et al., 1978) and DEA-BCC (Banker et al., 1984) methods. The former considers overall efficiency and assumes constant returns to scale (CRS). The latter introduces more details about the method, looks at pure technical efficiency, and assumes variable returns to scale (VRS). This study takes DEA-CCR as a performance assessing model and depicts the mathematical formulation as follows.

$$
\begin{aligned}
& Max \quad D_0 = \frac{\sum_{j=1}^{m} q_{j0} y_{j0}}{\sum_{i=1}^{n} p_{i0} x_{i0}} \\
& subject \quad to \\
& \frac{\sum_{j=1}^{m} q_{jh} y_{jh}}{\sum_{i=1}^{n} p_{ih} x_{ih}} \le 1, \; for \quad h = 1, \dots, k \\
& p_{ih} \ge 0 \quad , q_{jh} \ge 0, \; for \quad i = 1, \dots, n; \quad j = 1, \dots, m; \quad h = 1, \dots, k
\end{aligned} \tag{1}
$$

Here, $D_0$ denotes the efficiency score of the $o$ th DMU; $x_{ih}$ denotes the observed value of input $i$ for DMU $h$; $y_{jh}$ denotes the observed value of output $j$ for DMU $h$; $p_{ih}$ and $q_{jh}$ express the weighted attached to input $i$ and to output $j$ of DMU $h$, respectively; $k$ depicts the number of DMUs; $m$ describes the number of output variables; and $n$ represents the number of input variables.

Restructuring this formulation into linear form, the value of outputs/value of inputs $\leqq 1$ means that the value of outputs minus the value of inputs is less than or equal to 0.

$$
\begin{aligned}
& Max \quad D_0 = \sum_{j=1}^{m} q_{jo} y_{jo} \\
& subject \quad to \\
& \sum_{i=1}^{n} p_{io} y_{io} = 1 \\
& \sum_{j=1}^{m} q_{jh} y_{jh} - \sum_{i=1}^{n} p_{ih} y_{ih} \le 0, \; for \quad h = 1, \dots, k \\
& p_{ih} \ge 0, \quad q_{jh} \ge 0, \; for \quad i = 1, \dots, n; \quad j = 1, \dots, m, \quad h = 1, \dots, k
\end{aligned} \tag{2}
$$

### *Social network analysis: SNA*

SNA takes advantage of different types of approaches that can be used to clarify some properties about entities in the network (Parand et al., 2016). The concept of centrality was first introduced in connection with SNA, and the decided terminologies represent the sociological origin of such networks in a most suitable way (Bonacich, and Lloyd, 2001). Here, we introduce three different kinds of centralities: degree centrality, closeness centrality, and betweenness centrality. Degree centrality is used to depict the number of links incident upon a node. Equation (3) represents the degree centrality of a vertex $V$ for graph $G$.

$$C_D(V) = Deg(V) \tag{3}$$

Here, $V$ denotes the vertices, and $E$ denotes the edges.

Closeness centrality is defined as the sum of a node's distance from all other nodes. The mathematical formulation is expressed in Eq. (4).

$$C_C(x) = \frac{1}{\sum_y d(y,x)} \tag{4}$$

Bavelas (1950) stated that the more central a node is, the lower is its total distance from all other nodes.

$$H(x) = \sum_{y \neq x} \frac{1}{d(y,x)} \tag{5}$$

Betweenness centrality is utilized to quantify the number of occurrences of the node as a bridge along the shortest path between two nodes. Brandes (2001) indicated that the betweenness of a vertex $V$ in a graph $G := (V, E)$ can be decided by Eq. (6).

$$C_B(V) = \sum_{S \neq v \neq t \in V} \frac{\alpha_{st}(V)}{\alpha_{st}} \tag{6}$$

Here, $\alpha_{st}$ expresses the number of shortest paths from node $s$ to $t$; $\alpha_{st}(V_i)$ represents the number of shortest paths from node $s$ to $t$ that pass through $V_i$.

### *Core vector machine: CVM*

Support vector machine (SVM), a relative new machine learning method, has been widely adopted in numerous research domains due to its superior generalization ability. In practice, SVM utilization typically has some challenges in handling with large-scale datasets, such as tremendous training time and large space requirement (Collobert et al., 2002). Tsang et al. (2005) introduced core vector machine (CVM), which transforms the QP problem of SVM into the MEB problem so as to massively reduce the computational burdens. Experiment shows that CVM has comparable forecasting performance as does SVM, but is much faster and produces much fewer core vectors in dealing with large-scale datasets. To our knowledge, no existing research work has utilized CVM as a warning model for corporate operating performance forecasting. To fill this gap, this study employs CVM.

## Empirical results

### *The research samples*

The electronics industry in Taiwan not only has considerable influence on the global supply chain of consumer products, but the local stock market has also become an essential capital market for global investors as well. Moreover, the public authorities have announced numerous financial incentives or statutes for upgrading industries into this specific industry, turning it into an investment mainstream in the stock market. In fact, stocks in the electronics subsector typically make up over 70% of turnover transactions in Taiwan's stock market. Hence, this specific industry is our research target, and the data were collected from the Taiwan Economic Journal (TEJ) data bank covering 2014-2015.

### *The condition variables*

It is widely recognized that bad corporate operating performance is the main cause of financial distress and both are highly related. Thus, the condition variables used in financial distress prediction are taken as the surrogate for dealing with the task of operating performance forecasting. Table 1 presents the condition variables.

| **Measure** | | | | |
|---|---|---|---|---|
| Profitability | Liquidity | Leverage | Efficiency | Cash Position |
| C1: NI/TA | C3: CA/CL | C5: TL/TA | C7: COGS/I | C9: Cash/CL |
| C2: EBIT/TA | C4: WC/TA | C6: LTD/TA | C8: S/TA | C10: Cash/TA |
| NI: Net income; TA: Total assets; EBIT: Earnings before interest and tax; CA: Current assets; CL: Current liabilities; WC: Working capital; TL: Total liabilities; LTD: Long-term debts; COGS: Cost of goods sold; I: Inventory; S: Sales | | | | |

**Table 1. The Condition Variables**

### *Competitive edge determination by TM and SNA*

Yahoo!Finance, one of the largest global social media news providers, contains a large amount of information about corporates around the world. Business-related news about one corporate often mentions several other corporates in the same article. The corporate and any other mentioned corporates usually have some sort of business relation, such as a supply chain relationship or cooperative relationship. Bao et al. (2008) indicated that corporates appearing on the same webpage have a higher possibility of having some business relations. Thus, we assume that two corporates that are covered by the news many times imply that both of them have some business relations. We therefore make use of the stock tickers in business-related news provided by Yahoo!Finance and implement a text mining (TM) technique to construct the business relation network. After constructing the network, the study implements a social network analysis (SNA) technique to determine the corporate's competitive edge in this network and to examine its influence on the forecasting ask. Figure 1 illustrates a partial business relation network.

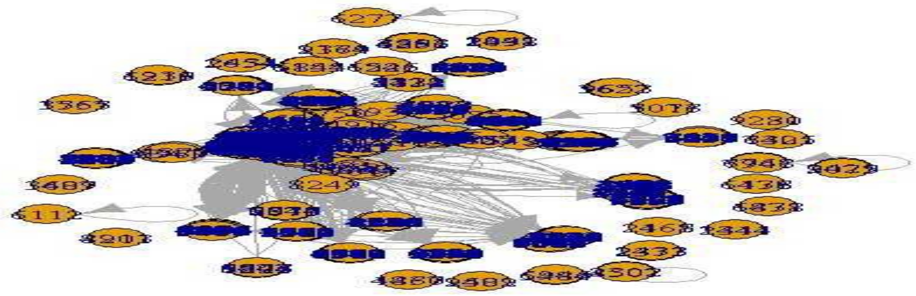

**Figure 1. The partial business relation network**

### *The decision variable*

Most previous studies on performance assessment mainly concentrate on monetary-based measures, such as ROA and ROE. However, in today's knowledge-based economic environment, the core element of a valued-creating business activity has shifted from monetary-based measures to non-monetary measures. In other words, a performance assessment restricted to monetary-based measures is unable to depict the full perspective of a corporate's operations. To deal with this obstacle, BSC (see Table 2) was initiated. Although BSC has an overarching consideration on performance assessment, it comes with a critical challenge of being unable to generate a final aggregated outcome. Thus, in order to handle the aforementioned task, this research utilizes DEA with its capabilities of dealing with multiple input and output variables simultaneously and providing an aggregated performance rank. By integrating BSC and DEA, a corporate's performance score can be decided. Corporates located in the highest ranking quintile (top 20%) of performance score are deemed as exhibiting superior operating performance, while corporates located in the lowest ranking quintile (bottom 20%) are deemed as showing inferior operating performance.

| **Perspectives** | | | |
|---|---|---|---|
| Financial | Customer | Internal Process | Learning and Innovation |
| F1: ROA | C1: MS | I1: ST | LI1: R&D |
| F2: OI | C2: CS | I2: NS | LI2: Tobin Q |
| F3: SG | C3: CR | I3: SE | LI3: HEP |
| F1: Return on assets; F2: Operating incomes; F3: Sales growth; C1: Market share; C2: Customer satisfaction; CE: Customer retention; I1: Sales to total assets; I2: Number of suppliers; I3: Sales to employees; LI1: Research and development expenditure; LI2: Tobin Q; LI3: High-educated proportion | | | |

**Table 2. The Four Perspectives in Balanced Scorecards**

### *Assessing criteria*

How to evaluate a model's forecasting quality is an essential topic for any task of forecasting. The accuracy or error rate is the most commonly adopted measure in financial risk classification, but utilizing one assessing measure to determine a model's forecasting performance is not reliable. Thus, this study further considers two other measures: type I error and type II error. Type I error means a corporate with superior operating performance is misclassified as a corporate with inferior operating performance; type II error represents a corporate with inferior operating performance being misclassified as a corporate with superior operating performance. Table 3 shows the mathematical formulations of assessing measures and depicts the confusion matrix.

Accuracy: (TN+TP)/(TP+FP+FN+TN)

Type I error: FP/(FP+TN)

Type II error: FN/(TP+FN)

| Condition | Predicted positive | Predicted negative |
|---|---|---|
| Actual positive | TP (True positive) | FN (False negative) (Type II error) |
| Actual negative | FP (False positive) (Type I error) | TN (True negative) |

**Table 3. The Confusion Matrix**

## *Forecasting results*

Feature selection aims to identify the most essential subset of features utilizing an assessing measure, which is an inevitable pre-process in forecasting model construction. It not only tremendously reduces computation complexity and storage cost, but also prevents the problem of over-fitting and biased outcome. The basic idea of the t-test method is used to determine whether there is a significant difference between two groups' means. It has been widely adopted to deal with the feature selection task in business and finance domains with acceptable performance.

This study thus employs the t-test method, and the selected features are represented in Table 4. We can see that a corporate with superior operating performance usually poses higher profitability, suitable capital structure, lower leverage, and good network position. One of the interesting findings is that a corporate's competitive edge (i.e., C11) is picked up as an essential feature. To examine its influence on performance forecasting, the experimental design is divided into two scenarios: with vs. without situations. To test the forecasting capability of our introduced model, we take it as benchmark and further compare it with three other forecasting models: neural network (NN), decision tree (DT), and Baye's network (BN). To prevent the result from happening by chance, this study adopts a non-parametric test - namely, the Wilcoxon signed-rank test.

Table 5 shows that adopting feature selection not only can increase a model's forecasting accuracy, but also can decrease both of the error types. Table 6 demonstrates the effectiveness of a corporate's competitive edge. This finding is in accordance with the previous work of Diez (2000) who indicated that a corporate that is highly involved in a business relation network can enlarge its profit margin by quickly responding to customer requirements with less uncertainty and facilitate its innovation capability by increasing its exposure to other valuable resources and information. Thus, a corporate's business relations can be viewed as reputational advantages and its competitive edge (Gnyawali and Madhavan, 2001).

| Measure | Variables (Selected ■ ; Non-selected □) |
|---|---|
| Profitability | ■C1: NI/TA; □C2: EBIT/TA |
| Liquidity | ■C3: CA/CL; □C4: WC/TA |
| Leverage | □C5: TL/TA; ■C6: LTD/TA |
| Efficiency | □C7: COGS/I; ■C8: S/TA |
| Cash position | □C9: Cash/CL; ■C10: Cash/TA |
| Competitive edge | ■C11: CE |

**Table 4. The Selected Attributes**

| Classifier | Measure | Scenario | | P-value |
|---|---|---|---|---|
| | | With FS | Without FS | |
| CVM | ACC | 84.85 | 79.05 | 0.039 |
| | Type I | 84.20 | 78.60 | 0.042 |
| | Type II | 85.50 | 79.50 | 0.039 |
| NN | ACC | 80.80 | 75.35 | 0.043 |
| | Type I | 79.80 | 72.20 | 0.043 |

| | | | | |
|---|---|---|---|---|
| | Type II | 81.80 | 78.50 | 0.039 |
| DT | ACC | 81.00 | 72.50 | 0.042 |
| | Type I | 81.50 | 70.00 | 0.043 |
| | Type II | 80.50 | 75.00 | 0.043 |
| BN | ACC | 80.55 | 75.25 | 0.043 |
| | Type I | 80.00 | 70.50 | 0.042 |
| | Type II | 81.10 | 80.00 | 0.273 |

**Table 5. The Results**

| Classifier | Measure | Scenario | | P-value |
|---|---|---|---|---|
| | | With C11: CE | Without C11: CE | |
| CVM | ACC | 84.85 | 79.05 | 0.042 |
| | Type I | 84.20 | 79.30 | 0.042 |
| | Type II | 85.50 | 78.80 | 0.043 |
| NN | ACC | 80.80 | 76.05 | 0.042 |
| | Type I | 79.80 | 75.10 | 0.039 |
| | Type II | 81.80 | 77.00 | 0.041 |
| DT | ACC | 81.00 | 76.25 | 0.043 |
| | Type I | 81.50 | 76.90 | 0.043 |
| | Type II | 80.50 | 75.60 | 0.042 |
| BN | ACC | 80.55 | 74.00 | 0.042 |
| | Type I | 80.00 | 73.00 | 0.042 |
| | Type II | 81.10 | 75.00 | 0.041 |

**Table 6. The Results**

## Conclusions

Corporate financial distress not only can cause large economic and social losses for market participants, but also can tremendously violate the stability of a stock market's development. Because of this, financial distress forecasting has become an attractive research topic in business and accounting domains over many decades. Kamei (1997) stated that up to 99% of financial distress cases are due to bad corporate operating performance - that is to say, bad corporate operating performance is an inevitable prior stage to financial distress. However, compared to well-studied research works on financial distress and credit risk forecasting, research works on corporate operating performance forecasting are quite rare.

To fill this gap in the literature, this study introduces a new model for corporate operating performance forecasting. Most previous studies on financial performance assessments are restricted to monetary-based measures, but utilizing them cannot reliably represent the whole facets of a corporate's operating condition. To overcome this obstacle, this study employs an overarching measure, called BSC, which contains monetary-based and non-monetary-based measures. The criticism on BSC is that it does not provide a synthesized final direction for decision makers to follow. To solve this problem, DEA was conducted. The performance rank was then decided by integrating BSC and DEA, and financial information derived from financial statements was then used to construct the forecasting model.

This study then incorporated TM and SNA to exploit corporate citations in social media news in order to establish business relation networks whose structural feature is implemented to infer a corporate's competitive edge. The result indicates that the model with the variable of competitive edge not only can increase its forecasting quality, but also can eliminate both of the error types. This finding agrees with Tsai (2001) who stated that the business relation network position of a corporate (i.e., a corporate's competitive edge) offers possibilities to learn from other corporates and to achieve innovative ideas that are essential to developing new products. The introduced model was examined by real-life cases and exhibits outstanding forecasting performance. Public authorities can consider the potential implications when formulating and announcing future policies. For example, they can give some financial incentives to or invest valuable resources into corporates with excellent business relations in order to facilitate their market competitiveness as well as to spur the prosperity of the related upstream and downstream industries. Through suitable investing strategies and policies, a country's industrial level can be steadily promoted and upgraded.

## Acknowledgement

We thank Metal Industrial Research & Development Centre (MIRDC) for their technical support. The authors would like to thank Ministry of Science and Technology of the Republic of China, Taiwan for financially supporting this work under Contract No. 105-2410-H-110 -033 -MY2.

## References

Acquaah, M., 2011. "Utilization and Value of Social Networking Relationships in Family and Nonfamily Firms in an African Transition Economy," European Management Journal (29:5), pp.347-361.

Altman, E. I. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," Journal of Finance (23:4), pp. 589-609.

Bao, S., Li, R., Yu, Y., and Cao, Y. 2008. "Competitor Mining with the Web," IEEE Transactions on Knowledge and Data Engineering (20:10), pp. 1297-1310.

Bavelas, A. 1950. "Communication Patterns in Task-oriented Groups," Journal of the Acoustical Society of America (22:6), pp. 725-730.

Banker, R. D., Charnes, A., and Cooper, W.W., 1984. "Some Models for the Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis," Management Science (30), pp. 1078-1092.

Banker, R. D., Chang, H., and Pizzini, M. J. 2004. "The Balanced Scorecard: Judgmental Effects of Performance Measures Linked to Strategy," Accounting Review (79:1), pp.1-23.

Banker, R. D., Chang, H., and Lee, S. Y. 2010. Differential Impact of Korean Banking System Reforms on Bank Productivity. Journal of Banking & Finance (34:7), pp. 1450-1460.

Beaver, W. H. 1966. "Financial Ratios as Predictors of Failure," Journal of Accounting Research (4), pp. 71-111.

Bentes, A. V., Carneiro, J., Silva, J. F., and Kimura, H. 2012. "Multidimensional Assessment of Organizational Performance: Integrating BSC and AHP," Journal of Business Research (65:12), pp. 1790-1799.

Bonacich, P., and Lloyd, P. 2001. "Eigenvector Like Measures of Centrality for Asymmetric Relations," Social Networks, pp. 191-201.

Brandes, U. 2001. "A Faster Algorithm for Betweenness Centrality," Journal of Mathematical Sociology (25), pp. 163-177.

Charnes, A., Cooper, W.W and Rhodes, E. 1978. "Measuring the Efficiency of DMUs," European Journal of Operational Research (2), pp. 429-444.

Chang, T. M., Sung, C. H., Hu, G. H., Hsu, M. F., and Lin, K. P. 2015. "Identifying Highly Potential Enterprises with Social Computing on Supply Chain Networks," IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), pp. 261-262.

Chang, T. M., and Hsu, M. F. 2016. "Integration of Incremental Filter-Wrapper Selection Strategy with Artificial Intelligence for Enterprise Risk Management," International Journal of Machine Learning and Cybernetics, doi:10.1007/s13042-016-0545-8.

Cook, W. D., Ruiz, J. L., Sirvent, I., and Zhu, J., 2017. "Within-group Common Benchmarking Using DEA," European Journal of Operational Research (256:3), pp. 901-910.

Collobert, R., Bengio, S., and Bengio, Y. 2002. "A Parallel Mixture of SVMs for Very Large Scale Problems," Neural Computation (14:5), pp. 1105–1114.

Diez, J. R. 2000. "The Importance of Public Research Institutes in Innovative Networks-Empirical Results from the Metropolitan Innovation Systems Barcelona Stockholm and Vienna", European Planning Studies (8), pp. 451-463.

Dimitras, A., Zanakis, S. and Zopounidis, C. 1996. "A survey of business failure with an emphasis on prediction methods and industrial application," European Journal of Operational Research (90), pp. 487-513.

Dyer, J.H. and Nobeoka, K., 2000. "Creating and Managing a High-performance Knowledge-sharing Network: The Toyota Case," Strategic Management Journal (21:3), pp. 345-367.

Fallahpour, S., Lakvan, E. N., and Zadeh, M. H. 2017. "Using an Ensemble Classifier Based on Sequential Floating Forward Selection for Financial Distress Prediction Problem," Journal of Retailing and Consumer Services (34), pp. 159-167.

Farrell, M. J. 1957. "The Measurement of Productive Efficiency," Journal of the Royal Statistical Society (120:3), pp. 253-281.

Ferguson, B. R. 2000. "Implementing Supply Chain Management," Production and Inventory Management Journal, pp. 64-67.

Gan, X., Duanmu, J., Wang, J., and Cong, W. 2013. "Anomaly Intrusion Detection based on PLS Feature Extraction and Core Vector Machine," Knowledge-Based Systems (40), pp. 1-6.

Geng, R., Bose, I., and Chen, X. 2015. "Prediction of Financial Distress: An Empirical Study of Listed Chinese Companies Using Data Mining," European Journal of Operational Research (241:1), 236-247.

Gnyawali, D. R., and Madhavan, R. 2001. "Cooperative Networks and Competitive Dynamics: A Structural Embeddedness Perspective. Academy of Management Review (26), pp. 431-445.

Ghalayini, A. M., and Noble, J. S. 1996. The Changing Basis of Performance Measurement," International Journal of Operations & Production Management (16: 8), pp.63-80.

Guo, C., and Zhu, J. 2017. "Non-cooperative Two-stage Network DEA Model: Linear vs. Parametric Linear," European Journal of Operational Research (258: 1), pp. 398-400.

Hafeez, K., Zhang Y. and Malak, N. 2002. "Core Capabilities for Sustainable Competitive Advantage: A Structured Methodology for Identifying Core Capabilities," IEEE Transactions on Engineering Management (49:1), pp. 28-35.

Hwang, S. N., and Chang, T. Y. 2003. "Using Data Envelopment Analysis to Measure Hotel Managerial Efficiency Change in Taiwan," Tourism Management (24:4), pp. 357-369.

Kamei, T., 1997. Risk Management (in Japanese). Tokyo: Dobunkan.

Kaplan, R. S., and Norton, D. P. (1992). "The Balance Scorecard – Measures that Drive Performance," Harvard Business Review (70:1), pp.71-79.

Kaplan, R. S. and Norton, D. P. 1996. "Using the Balanced Scorecard as a Strategic Management System," Harvard Business Review (74:1), PP. 75-85.

Kim, H. S., and Sohn, S. Y. 2010. "Support Vector Machines for Default Prediction of SMEs based on Technology Credit," European Journal of Operational Research (201:3), pp. 838-846.

Kim, D. Y. 2014. "Understanding Supplier Structural Embeddedness: A Social Network Perspective," Journal of Operations Management (32:5), pp. 219-231.

Kim, M. J., Lee, C. K., and Preis, M. W. 2016. "Seniors' Loyalty to Social Network Sites: Effects of Social Capital and Attachment," International Journal of Information Management (36:6), pp. 1020-1032.

Laitinen, E. K. and Laitinen, T. 2000. "Bankruptcy Prediction: Application of the Taylor's Expansion in Logistic Regression," International Review of Financial Analysis (9), pp. 327-349.

Li, H., Li, C. J., Wu, X. J., and Sun, J. 2014. "Statistics-based Wrapper for Feature Selection: An Implementation on Financial Distress Identification with Support Vector Machine," Applied Soft Computing (19), pp. 57-67.

Ma, Z., Pant, G., and Olivia Sheng, R. L. 2011. "Mining Competitor Relationships from Online News: A Network-based Approach," Electronic Commerce Research and Applications (10:4), pp. 418-427.

Meyer, P. A., and Pifer, H. W. 1970. "Prediction of Bank Failures," Journal of Finance 25(4), pp. 853-868.

Naudé, P., Zaefarian, G., Tavani, Z. N., Neghabi, S., and Zaefarian, R. 2014. "The Influence of Network Effects on SME Performance," Industrial Marketing Management (43:4), pp. 630-641.

Ohlson, J. A. 1980. "Financial Ratios and the Probabilistic Prediction of Bankruptcy," Journal of Accounting Research (18:1), pp. 109-131.

Parand, F. A., Rahimi, H., and Gorzin, M. 2016. "Combining Fuzzy Logic and Eigenvector Centrality Measure in Social Network Analysis," Physica A: Statistical Mechanics and its Applications (459), pp. 24-31.

Ponte, B., Sierra, E., Fuente, D., and Lozano, J. 2017. "Exploring the Interaction of Inventory Policies across the Supply Chain: An Agent-based Approach," Computers & Operations Research (78), pp. 335-348.

Qiao, P., Ju, X., and Fung, H. G. 2014. "Industry Association Networks, Innovations, and Firm Performance in Chinese Small and Medium-sized Enterprises," China Economic Review, (29), pp.213-228.

Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., and Ahuja, C. K. 2016. "A Package-SFERCB-Segmentation, Feature Extraction, Reduction and Classification Analysis by both SVM and ANN for Brain Tumors," Applied Soft Computing (47), pp. 151-167.

Sagarra, M., Mar-Molinero, C., and Agasisti, T. 2017. Exploring the Efficiency of Mexican Universities: Integrating Data Envelopment Analysis and Multidimensional Scaling. Omega (67), pp. 123-133.

Shafiee, M., Lotfi, F. H., and Saleh, H. 2014. "Supply Chain Performance Evaluation with Data Envelopment Analysis and Balanced Scorecard Approach," Applied Mathematical Modelling (38:21–22), pp. 5092-5112.

Sharma, V. K., Chandna, P., and Bhardwaj, A. 2017. "Green Supply Chain Management Related Performance Indicators in Agro Industry: A review," Journal of Cleaner Production (141), pp. 1194-1208.

Shen, Y. C., Chen, P. S., and Wang, C. H. 2016. "A study of enterprise resource planning (ERP) system performance measurement using the quantitative balanced scorecard approach," Computers in Industry (75), pp. 127-139.

Shirata, C. Y., and Sakagami, M., 2008. "An Analysis of the Going Concern Assumption": Text Mining from Japanese Financial Reports," Journal of Emerging Technologies in Accounting (5:1), pp. 1-16.

Tsai, W. 2001. "Knowledge Transfer in Intraorganizational Networks: Effects of Network Position and Absorptive Capacity on Business Unit Innovation and Performance," Academy of Management Journal (44), pp. 996-1004.

Tsang, I. W., Kwok, J. T., and Cheung, P. M. 2005. "Core Vector Machines: Fast SVM Training on Very Large Data Sets," Journal of Machine Learning Research (6), pp. 363–392.

Tjader, Y., May, J. H., Shang, J., Vargas, L. G., and Gao, N. 2014. "Firm-level Outsourcing Decision Making: A Balanced Scorecard-based Analytic Network Process Model," International Journal of Production Economics (147), pp. 614-623.

Wu, H. Y. 2012. "Constructing a Strategy Map for Banking Institutions with Key Performance Indicators of the Balanced Scorecard," Evaluation and Program Planning (35:3), pp. 303-320.

Wilson, R. L., and Sharda, R. 1994. "Bankruptcy Prediction Using Neural Networks," Decision Support Systems (11: 5), pp.545-557.

Xue, X., Wu, H., Zhang, X., Dai, J., and Su, C. 2015. "Measuring Energy Consumption Efficiency of the Construction Industry: the case of China," Journal of Cleaner Production (107), pp. 509-515.

Xia, D., Yu, Q., Gao, Q., and Cheng, G. 2017. "Sustainable Technology Selection Decision-Making Model for Enterprise in Supply Chain: Based on a Modified Strategic Balanced Scorecard," Journal of Cleaner Production (141), pp. 1337-1348.

Yin, S., and Yin, J., 2016. "Tuning Kernel Parameters for SVM Based on Expected Square Distance Ratio," Information Sciences (370–371), pp. 92-102.

Zhou, M., Dan, B., Ma, S., and Zhang, X. 2017. "Supply Chain Coordination with Information Sharing: The Informational Advantage of GPOs," European Journal of Operational Research (256:3), pp. 785-802.