Computational Economics
https://doi.org/10.1007/s10614-021-10115-8

A New Strategy for Short-Term Stock Investment Using Bayesian Approach

Tai Vo-Van[1] · Ha Che-Ngoc[2,3] · Nghiep Le-Dai[4] · Thao Nguyen-Trang[2,3]

Accepted: 18 March 2021

Abstract
In this paper, an application of the Bayesian classifier for short-term stock trend prediction is presented. In order to use Bayesian classifier effectively, we transform the daily stock price time series object into a data frame format where the dependent variable is the stock trend label and the independent variables are the stock variations of the last few days. Based on the posterior probability density function, we propose a new method for stock selection and then propose a new stock trading strategy. The numerical examples demonstrate the potential of the proposed strategy for application to short-term stock trading.

Keywords Stock prediction · Stock selection · Bayesian classifier · One-step prediction · Two-step prediction · Bayes error

1 Introduction

Recently, along with the increasing number of joint-stock companies, the stock market has become more and more vibrant; and therefore, stock investing has been a hot research topic (Chen et al. 2014; Chen 2008; Mladjenovic 2016). In general, there are two major stock investing strategies consisting of technical analysis and fundamental analysis (Quah 2008). Fundamental analysis is mainly used for long-term investment by checking a company's financial features, such as average equity, average asset, sales cost, revenues, operating profit, and net income, etc (Hajjami

✉ Thao Nguyen-Trang
nguyentrangthao@tdtu.edu.vn

[1] College of Natural Science, Can Tho University, Can Tho City, Vietnam

[2] Division of Computational Mathematics and Engineering, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[3] Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[4] College of Basis Science, Nam Can Tho University, Can Tho, Vietnam

Published online: 03 April 2021

and Amin 2018; Pätäri et al. 2018; Zhai and Bai 2018). Some of the recent fundamental analysis strategies include the mean-variance model (Markowitz 1952), the data envelopment analysis (Chen 2008; Huang et al. 2015; Zhou et al. 2018), and the ordered weighted averaging operator (Amin and Hajjami 2016; Hajjami and Amin 2018). Long-term investment is encouraged for investors because it can create a sustainable business; however, it takes a long time for investors to generate profit. In addition to fundamental analysis, investors are also interested in technical analysis to get short-term profit (Quah 2008). Instead of analyzing the financial statements, technical analysis focuses more on historical price movement. Technical analysis is interested in studying historical data to predict the future. There is a broad range of technical analysis studies, both simple methods, such as chart analysis (Elliott 2007; Patel 2010; Roscoe and Howorth 2009; Vasiliou et al. 2006), and complex methods, such as machine learning approach (Alfarano et al. 2005; Atsalakis et al. 2016; Batra and Daudpota 2018; Deng et al. 2015; Ghasemiyeh et al. 2017; Gupta and Wang 2010; Hu et al. 2018; Huarng and Yu 2005; Jeon et al. 2018; Kale et al. 2018; Kohli et al. 2019; Liu and Zhang 2019; Maciel and Ballini 2020; Parmar et al. 2018; Sang et al. 2019; Sollis et al. 2000; Syriopoulos et al. 2020; Usmani et al. 2016; Wiesinger et al. 2013; Xu and Cohen 2018; Yang et al. 2019; Zhang 2003; Zhang and Tan 2018). Technical analysis methods are diverse and varied, but in general, most of them aim to identify peaks and troughs so that investors can "buy at the low and sell at the high" and get the highest expected return (Cartea et al. 2014; Goldman et al. 2018; Zervos et al. 2012).

In short-term investment, it is more important to accurately predict the increase or decrease of stock prices than to accurately predict the exact stock value. As shown in Fig. 1, the orange line is the actual stock values, the green line and the blue line are the predictions of Method 1 and Method 2, respectively. Compared to the actual value, Method 1 results in an error of 2, whereas Method 2 results in an

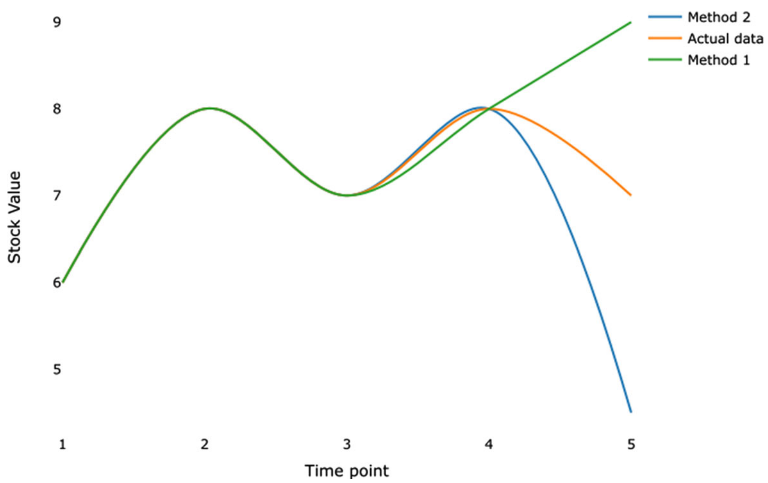


Fig. 1 The prediction of the two methods

error of 2.5. Based on the error values, investors may believe that Method 1 is better than Method 2, but this can lead to serious mistakes. In fact, Method 1 gives a lower error than Method 2 but it completely mispredicted the trend of the stock. Using Method 1, investors might still hold on to the stock at the time t and expect further up-move. However, the stock market peak occurred at the time t and fell at time $t+1$, which leads to a loss. For Method 2, although it results in lower performance in terms of forecasting the stock value, it is capable of capturing the stock price trend, from which the stock might be sold at the peak. Thus, it can be believed that accurately predicting the stock trend is more important than approximating the stock price.

In order to accurately predict the stock trend, we need to compute the variations or the first order differences of the stock values rather than the original stock values. As shown in Fig. 2, when the current stock price is 1, the stock price in the next time can rise and fall, arbitrarily. In contrast, if we are interested in the fluctuation of n days before the forecast time, some interesting rules can be discovered. For example, as shown in Fig. 2, if the stock price fell in the two previous days (the first order difference is less than 0), the stock price will rise in the current day; also, if the stock price rose the two previous days, the stock price will fall in the current day. The mentioned rules are also consistent with which we believe that when the stock price has fallen/risen for a few days, it will find the support/resistance and reverse. In fact, the found rules will be more complex and also contains uncertainty.

According to the above discussion, firstly, this paper introduces a method to predict the short-term stock trend based on the first order difference of stock price. Specifically, the independent variables are the first order differences of stock prices of n days before the forecast time and the binary dependent variable represents the rise/fall of the stock. For this purpose, the time series collected in the past would be transformed into a data frame and then would be trained by a supervised learning model. In this paper, we use the Bayesian classifier because it not only can classify

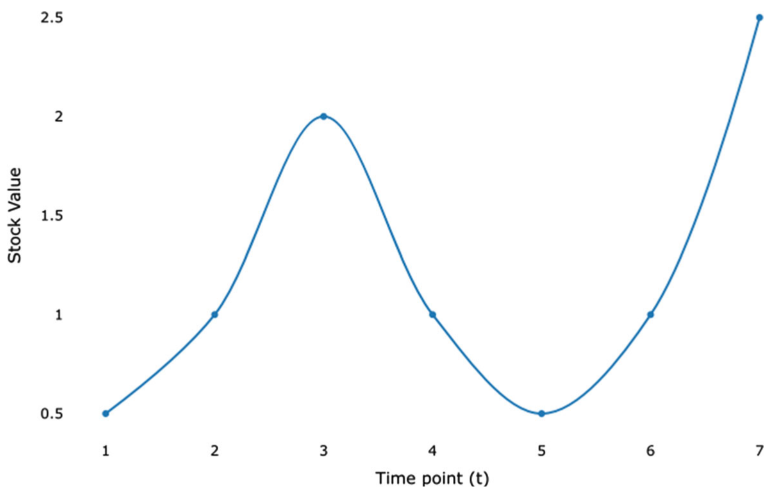


Fig. 2 A time series of stock

Table 1 The contributions of the proposed method compared to recent methods in literature

	Independent variables	Dependent variable	Prediction	Stock selection method
Batra and Daudpota (2018)	News	Stock trend	One step	–
Hu et al. (2018)	News	Stock trend	One step	–
Jeon et al. (2018)	Historical price	Stock price	One step	
Sang et al. (2019)	–	–	–	Fundamental features
Xu and Cohen (2018)	New and historical price	Stock trend	One step	–
Yang et al. (2019)	Historical price	Stock price	One step	Stock price and Fundamental features
Zhang and Tan (2018)	Historical price	Stock price	One step	Stock price
Proposed method	Historical variations	Stock trend	One & two steps	Bayes error

the data but also provides the predictive probability of classification, which helps us can evaluate the reliability of the predicted result (Addesso et al. 2013; Castellaro et al. 2017; Nguyen-Trang and Vo-Van 2017; Pizzo et al. 2018; Vovan 2017). Secondly, because of the large number of companies on the stock market, a model for stock selection needs to be developed. Most of current stock selection models aim at maximizing expected return based on the predicted stock price and fundamental features. However, a maximum expected return in fact only occurs when peaks and troughs are correctly identified. In this paper, we propose a stock selection method based on maximizing probability of a correct classification or minimizing the Bayes error, leading to a higher probability of a correct identifying of peaks and troughs. In other words, instead of maximizing expected return, the proposed stock selection method aims at maximizing probability of a correct identifying of peaks and troughs, which limits a lot of risk and ensures a relatively high profit. Finally and most importantly, for practical investigating, in addition to one-step ahead prediction, we also propose a two-step ahead prediction method and propose a strategy of decision-making. The contributions of this paper are summarized in Table 1.

The remainder of the paper is presented as follows. Section 2 theoretically reviews the Bayesian classifier and the Bayes error. Section 3 presents the proposed method. Section 4 presents a few experimental results on real data sets. The final section is destined for the conclusion.

2 Bayesian Classifier and Bayes Error

2.1 Bayesian Classifier

We consider k classes, $w_1, w_2, \ldots, w_k$, with the prior probability q_i, $i = \overline{1,k}$. $\mathbf{X} = \{X_1, X_2 \ldots, X_n\}$ is the n-dimensional continuous data with $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ is a specific sample. According to Nguyen-Trang and Vo-Van (2017), Pham-Gia et al. (2008), a new observation $\mathbf{x}$ belongs to the class w_i if

$$P(w_i|\mathbf{x}) > P\left(w_j|\mathbf{x}\right) \quad \text{for} \quad 1 \le j \le k, j \ne i. \tag{1}$$

In continuous case, $P(w_i|\mathbf{x})$ is calculated by:

$$P(w_i|\mathbf{x}) = \frac{P(w_i)f(\mathbf{x}|w_i)}{\sum_{i=1}^{k} P(w_i)f(\mathbf{x}|w_i)} = \frac{q_i f_i(\mathbf{x})}{f(\mathbf{x})}. \tag{2}$$

Because $f(\mathbf{x})$ is the same for all classes, the classification's rule is:

$$q_i f_i(\mathbf{x}) > q_j f_j(\mathbf{x}) \Leftrightarrow \frac{f_i(\mathbf{x})}{f_j(\mathbf{x})} > \frac{q_j}{q_i}, \forall i \ne j, \tag{3}$$

where q_i and $f_i(\mathbf{x})$ are the prior probability and the probability density function of class i, respectively.

In case of two classes like the stock trend prediction problem, the a new observation $\mathbf{x}$ belongs to the class w_1 if $q_1 f_1(\mathbf{x}) > q_2 f_2(\mathbf{x})$ or $P(w_1|\mathbf{x}) > 0.5$ and vice versa.

2.2 Bayes Error

Let $g_i(\mathbf{x}) = q_i f_i(\mathbf{x})$, $g_{\max}(\mathbf{x}) = \max_i\{g_i(\mathbf{x})\}$ and $g_{\min}(\mathbf{x}) = \min_i\{g_i(\mathbf{x})\}$. As presented in Nguyen-Trang and Vo-Van (2017), Pham-Gia et al. (2008), Bayes error is calculated by Formula 4.

$$Pe_{1,2,\ldots,k}^{(q)} = 1 - \int_{R^n} g_{\max}(\mathbf{x}) d\mathbf{x} \tag{4}$$

where Pe is Bayes error, n is the dimension of the data.

Theorem 1 *In the general k-class Bayesian classifier, we have the following results on Bayes error*:

1.

$$0 \le Pe_{1,2,\ldots,k}^{(q)} \le 1 - \max_i\{q_i\}, k \ge 2. \tag{5}$$

2. $\frac{1}{k}\left(k - 1 - \sum_{i<j} \int_{R^n} |g_i(\mathbf{x}) - g_j(\mathbf{x})| d\mathbf{x}\right)$

$$\le Pe_{1,\ldots,k}^{(q)} \le \frac{1}{2} \max_{i<j}\left\{\int_{R^n} |g_i(\mathbf{x}) - g_j(\mathbf{x})| d\mathbf{x}\right\} + \min_i\{q_i\}. \tag{6}$$

Proof

1. With each $i = 1, 2, \ldots, k$, we have
   $q_i f_i(\mathbf{x}) \le \max_i\{q_i f_i(\mathbf{x})\} \le \sum_{i=1}^{k} q_i f_i(\mathbf{x})$ and $\int_{R^n} f_i(\mathbf{x}) d\mathbf{x} = 1$.
   After integrating the above inequality, we obtain:

$$q_i \le \int_{R^n} g_{\max}(\mathbf{x}) d\mathbf{x} \le 1.$$

   Thus,

$$\max_i\{q_i\} \le \int_{R^n} g_{\max}(\mathbf{x}) d\mathbf{x} \le 1,$$

   or

$$0 \le 1 - \int_{R^n} g_{\max}(\mathbf{x}) d\mathbf{x} \le 1 - \max_i\{q_i\}.$$

   That means we have (5).
2. We have

$$\begin{aligned}
\int_{R^n} g_{\max}(\mathbf{x}) d\mathbf{x} \ge \max_{i<j} \int_{R^n} & \max\{g_i(\mathbf{x}), g_j(\mathbf{x})\} d\mathbf{x} \\
&= \max_{i<j}\left\{\frac{1}{2}\int_{R^n} |g_i(\mathbf{x}) - g_j(\mathbf{x})| d\mathbf{x} + \frac{1}{2}(q_i + q_j)\right\} \\
&\ge \max_{i<j}\left\{\frac{1}{2}\int_{R^n} |g_i(\mathbf{x}) - g_j(\mathbf{x})| d\mathbf{x}\right\} + \min_{i<j}\left\{\frac{1}{2}(q_i + q_j)\right\} \\
&\ge \max_{i<j}\left\{\frac{1}{2}\int_{R^n} |g_i(\mathbf{x}) - g_j(\mathbf{x})| d\mathbf{x}\right\} + \min_i\{q_i\}.
\end{aligned}$$

   Thus

$$Pe^{(q)}_{1,2,\ldots,k} \le \frac{1}{2}\max_i\left\{\int_{R^n} |g_i(\mathbf{x}) - g_j(\mathbf{x})| d\mathbf{x}\right\} + \min_i\{q_i\}. \tag{7}$$

   Besides, due to

$$\sum_{i<j}\sum_{j}|g_i(\mathbf{x})-g_j(\mathbf{x})| \geq \sum_{j=1}^{k}\left[\max_i\{g_i(\mathbf{x})\}-g_j(\mathbf{x})\right]$$
$$=k\left[\max_i\{g_i(\mathbf{x})\}\right]-\sum_{j=1}^{k}g_j(\mathbf{x}).$$

We have

$$\max_i\{g_i(\mathbf{x})\} \leq \frac{1}{k}\sum_{i<j}\sum_{j}|g_i(\mathbf{x})-g_j(\mathbf{x})|+\frac{1}{k}\sum_{j=1}^{k}g_j(\mathbf{x}).$$

Since

$$\sum_{j=1}^{k}\int_{R^n}g_j(\mathbf{x})\mathrm{d}\mathbf{x}=1$$

the above inequality becomes

$$\int_{R^n}g_{\max}(\mathbf{x})\mathrm{d}\mathbf{x} \leq \frac{1}{k}\sum_{i<j}\sum_{j}\int_{R^n}|g_i(\mathbf{x})-g_j(\mathbf{x})|\mathrm{d}\mathbf{x}+\frac{1}{k}. \tag{8}$$

From (8), we also obtain

$$Pe_{1,2,\ldots,k}^{(q)} \geq \frac{1}{k}\left(k-1-\sum_{i<j}\int_{R^n}|g_i(\mathbf{x})-g_j(\mathbf{x})|\mathrm{d}\mathbf{x}\right). \tag{9}$$

From (7) and (9), we obtain (6).

□

Theorem 2 *The relationship between Bayes error and Toussaint Distance.*

1.

$$Pe_{1,2,\ldots,k}^{(q)} \leq \sum_{i<j}q_i^{\beta}q_j^{1-\beta}D_T\left(f_i,f_j\right)^{(\beta,1-\beta)}$$

2.

$$Pe_{1,2,\ldots,k}^{(q)} \leq 1-\frac{1}{k-1}\left(1-\prod_{i=1}^{k}q_j^{\alpha_j}D_T(f_1,f_2,\ldots,f_k)^{\alpha}\right)$$

where $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_k)$, $\alpha_1+\alpha_2+\ldots+\alpha_k=1$, *and* $D_T(f_1,f_2,\ldots,f_k)^{\alpha}=\int_{R^n}\{[f_1(\mathbf{x})]^{\alpha_1}\ldots[f_k(\mathbf{x})]^{\alpha_k}\}\mathrm{d}\mathbf{x}$ *is Toussaint Distance.*

Proof

1.

$$\begin{aligned} Pe^{(q)}_{1,2,\ldots,k} &= 1 - \int\limits_{R^n} g_{\max}(\mathbf{x})\mathrm{d}\mathbf{x} \\ &= 1 - \int\limits_{R^n} \max_{1\le l\le k}\{q_l f_l(\mathbf{x})\}\mathrm{d}\mathbf{x} \\ &= \int\limits_{R^n} \sum_{j=1}^{k} q_j f_j(\mathbf{x})\mathrm{d}\mathbf{x} - \sum_{j=1}^{k} \int\limits_{R^n_j} \max_{1\le l\le k}\{q_l f_l(\mathbf{x})\}\mathrm{d}\mathbf{x} \\ &= \sum_{j=1}^{k}\left[\int\limits_{R^n} q_j f_j(\mathbf{x})\mathrm{d}\mathbf{x} - \int\limits_{R^n_j} \max_{1\le l\le k}\{q_l f_l(\mathbf{x})\}\mathrm{d}\mathbf{x}\right] \\ &= \sum_{j=1}^{k} \int\limits_{R^n\setminus R^n_j} q_j f_j(\mathbf{x})\mathrm{d}\mathbf{x} \\ &= \sum_{j=1, j\neq i}^{k} \int\limits_{R^n_j} \min\{q_i f_i(\mathbf{x}), q_j f_j(\mathbf{x})\}\mathrm{d}\mathbf{x} \\ &= \sum_{i<j} \int\limits_{R^n_i} \min\{q_i f_i(\mathbf{x}), q_j f_j(\mathbf{x})\}\mathrm{d}\mathbf{x}. \end{aligned}$$

Since

$$\left[\min\{q_i f_i(\mathbf{x}), q_j f_j(\mathbf{x})\}\right]^{\beta} \le (q_i f_i(\mathbf{x}))^{\beta}$$

and

$$\left[\min\{q_i f_i(\mathbf{x}), q_j f_j(\mathbf{x})\}\right]^{1-\beta} \le (q_i f_i(\mathbf{x}))^{1-\beta}$$

then

$$\min\{q_i f_i(\mathbf{x}), q_j f_j(\mathbf{x})\} \le (q_i f_i(\mathbf{x}))^{\beta}(q_i f_i(\mathbf{x}))^{1-\beta}.$$

Integrating the above inequality, we obtain:

$$\begin{aligned} Pe^{(q)}_{1,2,\ldots,k} &\le \sum_{i<j} \int_{R^n_i} [(q_i f_i(\mathbf{x}))^{\beta}(q_j f_j(\mathbf{x}))^{1-\beta}]\mathrm{d}\mathbf{x} \\ &\le \sum_{i<j} \int_{R^n} [(q_i f_i(\mathbf{x}))^{\beta}(q_j f_j(\mathbf{x}))^{1-\beta}]\mathrm{d}\mathbf{x} \\ &= \sum_{i<j} q_i^{\beta} q_j^{1-\beta} D_T(f_i, f_j)^{(\beta, 1-\beta)}. \end{aligned}$$

2. For each $j = 1, 2, \ldots, k$, we have

$$\left(\sum_{j=1}^{k} q_j f_j(\mathbf{x})\right)^{\alpha_i} \geq (q_i f_i(\mathbf{x}))^{\alpha_i}, \quad i = 1, 2, \ldots, k.$$

Therefore,

$$\left(\sum_{j=1}^{k} q_j f_j(\mathbf{x})\right)^{\alpha_1+\alpha_2+\cdots+\alpha_k} \geq \prod_{j=}^{k} \left(q_j f_j(\mathbf{x})\right)^{\alpha_j} \Leftrightarrow \sum_{j=1}^{k} q_j f_j(\mathbf{x}) \geq \prod_{j=}^{k} \left(q_j f_j(\mathbf{x})\right)^{\alpha_j}. \tag{10}$$

On the other hand,

$$\left(\min_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\}\right)^{\alpha_1} \leq (q_1 f_1(\mathbf{x}))^{\alpha_1}, \ldots, \left(\min_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\}\right)^{\alpha_k} \leq (q_k f_k(\mathbf{x}))^{\alpha_k}.$$

So

$$\left(\min_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\}\right)^{\alpha_1+\cdots+\alpha_k} \leq \prod_{j=1}^{k} \left(q_j f_j(\mathbf{x})\right)^{\alpha_j},$$

or

$$\min_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\} \leq \prod_{j=1}^{k} \left(q_j f_j(\mathbf{x})\right)^{\alpha_j}. \tag{11}$$

Combining (10) and (11), we obtain

$$0 \leq \sum_{j=1}^{k} q_j f_j(\mathbf{x}) - \prod_{j=1}^{k} \left(q_j f_j(\mathbf{x})\right)^{\alpha_j} \leq \sum_{j=1}^{k} q_j f_j(\mathbf{x}) - \min_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\}.$$

Because $\sum_{j=1}^{k} q_j f_j(\mathbf{x}) - \min_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\}$ includes $(k-1)$ terms, we have

$$\sum_{j=1}^{k} q_j f_j(\mathbf{x}) - \min_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\} \leq (k-1) \max_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\}.$$

Thus

$$0 \leq \sum_{j=1}^{k} q_j f_j(\mathbf{x}) - \prod_{j=1}^{k} \left(q_j f_j(\mathbf{x})\right)^{\alpha_j} \leq (k-1) \max_{1 \leq j \leq k}\{q_j f_j(\mathbf{x})\}.$$

Integrating the above relation, we obtain:

$$1 - \prod_{j=1}^{k} q_j^{\alpha_j} D_T(f_1, f_2, \ldots, f_k)^{\alpha} \leq (k-1) \int_{R^n} g_{\max}(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{12}$$

Using $\int_{R^n} g_{\max}(\mathbf{x}) = 1 - Pe_{1,2,\ldots,k}^{(q)}$ for (12), Theorem 2 has been proved.

□

2.3 Bayes Error in Binary Classification

In case of binary classification,

$$Pe = 1 - \int_{R^n} g_{\max}(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{R^n} g_{\min}(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{13}$$

Straightforwardly, the Bayes error calculated using Formula 13 is the area of the $g_{\min}$ function or the area of the overlap region between the two classes. Figure 3 illustrates a case study having one independent variable. Assuming that we have two types consisting of w_1 (stock will rise), and w_2 (stock will fall), and $\mathbf{x}$ is the variation of the previous day. The probability of a false prediction is calculated by:

$$Pe = Pe_1 + Pe_2 \tag{14}$$

where Pe_1 is the probability that the predicted class is 1 but the actual class is 2, and Pe_2 is the probability that the predicted class is 2 but the actual class is 1. Figure 3 shows that

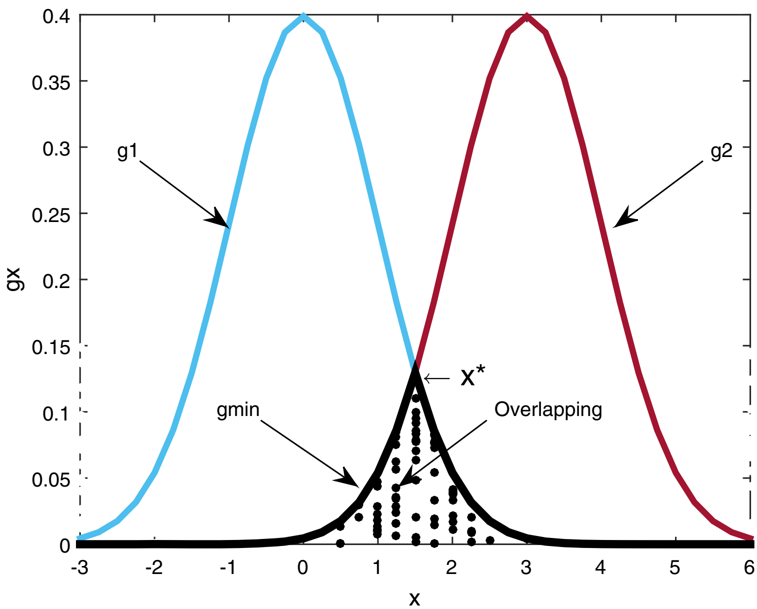


Fig. 3 Bayes error in case of binary classification

$$Pe_1 = \int_{-\infty}^{\mathbf{x}*} g_2(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{-\infty}^{\mathbf{x}*} g_{\min}(\mathbf{x})\mathrm{d}\mathbf{x} \tag{15}$$

and

$$Pe_2 = \int_{\mathbf{x}*}^{+\infty} g_1(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{\mathbf{x}*}^{+\infty} g_{\min}(\mathbf{x})\mathrm{d}\mathbf{x} \tag{16}$$

with $\mathbf{x}*$ is the root of equation $g_1(\mathbf{x}) = g_2(\mathbf{x})$.

Combining (14), (15) and (16), we obtain:

$$Pe = \int_{-\infty}^{+\infty} g_{\min}(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{17}$$

In case of univariate normal distribution, we can find out the specific expression for $g_{\min}(\mathbf{x})$ (Nguyen-Trang and Vo-Van 2017; Vovan 2017); hence, the *Pe* value can be identified. In the case of arbitrary multivariate distributions, a specific expression of $g_{\min}$ is difficult to be identified; hence, Quasi Monte-Carlo method is applied to approximate the value of integrals.

The Bayes error calculated by Formula (17) can be visually interpreted as the area of overlapping region between g_1 and g_2. Using the Bayes error, we can estimate the probability of an incorrect prediction, therefore indicating which stocks are easier to correctly predict. This property is fully applicable to portfolio selection. In this case, stocks are selected according to their corresponding Bayes error, with the lowest yielding more chances to be investigated. Stocks of this nature are easier to correctly predict, and safer than stocks with high profitability, but low probability of correct prediction.

3 Proposed Method

This section proposes a new method using the variations of the last two days to predict next day's stock trend. Firstly, in order to use the Bayesian classifier effectively, we need to transfer the time series data of stock price to tabular data including one dependent variable and two independent variables. This algorithm is presented in Sect. 3.1. In Sect. 3.2, we apply Bayesian classifier to training dataset before making one-step and two-step prediction. A short-term investment strategy based on predicted results is also constructed in this subsection. Section 3.3 presents a stock selection method to ensure high probability of correct prediction.

3.1 Transform the Time Series Data to Tabular Data

Algorithm 1 Given historical data $X(t)$, $t = 1, 2, \ldots, N$, with $x(t)$ is the specific value of $X(t)$ at time t, N is the length of the original time series, Algorithm 1 transforms the time series data to tabular data.

INPUT:	$x(t)$
FOR	$t = 1 : N$
	Compute the variation or first order difference:
	$v(t) := x(t) - x(t-1)$
ENDFOR	
FOR	$t = 3 : N$
	IF $v(t+1) > 0$
	$Y(t) := 1$
	ELSE
	$Y(t) := 0$
	ENDIF
ENDFOR	
	TrainingData:= $[v(t-1), v(t), \ldots, Y(t)], t = 3 : N$
OUTPUT:	TrainingData

3.2 Predict the Stock Trend

3.2.1 One-step Ahead Prediction

Using the Bayesian classifier on the training data obtained, we can predict the stock trend given the variations of the previous two days. This method is briefly presented in Algorithm 2.

Algorithm 2 Given training data, and the variations of the previous two days, this algorithm identifies whether the stock will rise (w_1) or not (w_2).

INPUT:	Training data, and the variations of the previous two days v_{t-1}, v_t.
	Build the Bayesian classifier.
	Compute $P(w_1 \vert v_{t-1}, v_t)$
IF	$P(w_1 \vert v_{t-1}, v_t) > 0.5$
	The stock will rise at time $t+1$.
ELSE	
	The stock will fall at time $t+1$.
ENDIF	
OUTPUT:	Class of the stock trend.

3.2.2 Two-step Ahead Prediction

The Algorithm 2 can only predict one-step ahead. For two-step ahead, namely the prediction of $t+2$, only the 2-lag (the variation at time t) is given. In other words, we need to calculate $P(w_1|v_{t-1}, v_t)$ with an unknown value of v_t. Let v_{t-1}^0 be the given value of v_{t-1}, $P(w_1|v_{t-1} = v_{t-1}^0, v_t)$ is calculated as follows:

$$P(w_1|v_{t-1} = v_{t-1}^0, v_t) = \frac{P\left(v_{t-1} = v_{t-1}^0, v_t|w_1\right)P(w_1)}{P\left(v_{t-1} = v_{t-1}^0, v_t\right)} = \frac{g_1\left(v_{t-1} = v_{t-1}^0, v_t\right)}{f\left(v_{t-1} = v_{t-1}^0, v_t\right)} \quad (18)$$

Similar to one-step ahead prediction, a new observation x belongs to the class w_1 if and only if $g_1\left(v_{t-1} = v_{t-1}^0, v_t\right) > g_2\left(v_{t-1} = v_{t-1}^0, v_t\right)$ or $P(w_1|v_{t-1} = v_{t-1}^0, v_t) > 0.5$ and vice versa. However, in this case, v_t is an unknown value of the random variable V_t; hence, $P(w_1|v_{t-1} = v_{t-1}^0, v_t)$ calculated according to Formula (18) is not a specific number but a random variable. Let Z be the random variable representing $P(w_1|v_{t-1} = v_{t-1}^0, v_t)$, using the law of the unconscious statistician, the expectation of Z can be obtained as follows.

$$E(Z) = \int_{-\infty}^{+\infty} \frac{g_1\left(v_{t-1} = v_{t-1}^0, v_t\right)}{f\left(v_{t-1} = v_{t-1}^0, v_t\right)} f(v_t)\mathrm{d}v_t \quad (19)$$

where $f(v_t)$ is the marginal probability density function of V_t. Formula (19) can be easily calculated with the support of statistical software; therefore, we can use $E(Z)$ to predict the stock trend at time $t+2$. If $E(Z) > 0.5$, we predict the stock will rise, and vice versa. The method of two-step ahead prediction is summarized in Algorithm 3.

Algorithm 3 Given training data, this algorithm predicts the stock trend at time $t+2$.

INPUT:	Training data
	Calculate $E(Z)$ using Formula (19).
IF	$E(Z) > 0.5$
	The stock will rise at time $t+2$.
ELSE	
	The stock will fall at time $t+2$.
ENDIF	
OUTPUT:	Class of the stock trend.

Table 2 Strategy of decision-making at time $t+1$

$t+1$	$t+2$	Decision
1	1	Wait
1	0	Sell
0	1	Buy
0	0	Wait

3.2.3 Strategy of Decision-Making

Based on the stock trend predicted at time $t+1$ and $t+2$, we propose a strategy of decision making at time $t+1$. Four scenarios are presented as follows and summarized in Table 2.

Scenario 1: We predict that the stock will rise at time $t+1$ and $t+2$. In that case, no peak is identified at time $t+1$; hence, investors should not make decision and wait for next signals.

Scenario 2: We predict that the stock will rise at time $t+1$ and fall at time $t+2$. In that case, a peak occurs at time $t+1$, investors should sell the stock at that moment.

Scenario 3: We predict that the stock will fall at time $t+1$ and rise at time $t+2$. In that case, a trough occurs at time $t+1$, investors should buy the stock at that moment.

Scenario 4: We predict that the stock will fall at time $t+1$ and $t+2$. In that case, no trough is identified at time $t+1$; hence, investors should not make decision and wait for next signals.

3.3 Stock Selection

Sections 3.1 and 3.2 have presented the methods for predicting stock trend and making trading decision. However, all models including the proposed model are wrong, not useful, and can make serious error when applied to unsuitable data. In other words, if the stock is not well selected, the probability of error is very high. In Sect. 2.2, Bayes error is an effective measure for evaluating the overlap degree of data as well as the probability of a false prediction. Therefore, only the stock with the lowest Bayes error can be selected. Let us have a list of n stocks and Pe_i be the Bayes error relevant to the ith stock. The jth stock can be selected if:

$$j = \arg \min_{i=\overline{1,n}} Pe_i.$$

In addition to the above formula, an error threshold can be set for multiple stock selections. Also, users can use a combination of Bayes error and other measures to fit their specific purposes.

4 Numerical Examples

In this section, four examples are presented to illustrate and test whether the proposed strategy is appropriate for applying in Vietnam Stock Market. In particular, Example 1 illustrates the short-term stock selection method, which has been presented in Sect. 3.3. Example 2 evaluates the proposed method performance in terms of the profits obtained. For comparison purposes, other approaches consisting of a no-skill algorithm, buy and hold strategy, ARIMA model and Ichimoku cloud charts are also performed and evaluated. Example 3 evaluates the stock selection method by comparing the profit obtained from various stocks. In Example 4, the performance of all comparative strategies is investigated in an extended investment period. In addition, for ensuring the quality of the experiments, only stocks within VN30 basket, which have high liquidity, are taken in the experiments.

4.1 Rolling Forward Method

For the Bayesian classifier and the ARIMA model, which need to be trained using the historical data, we use the rolling forward method rather than using a fixed training data set. According to this method, for predicting the stock value at time point t, only the stock price from time point $t - N$ to time point $t - 1$ are used for training purpose. Through the training process, this method not only can continuously update the newest data but also can remove the oldest data that have weak effect on the current time point. Due to this advantage, in the field of stock prediction, the rolling forward method was widely incorporated into the training process (Arratia 2014; Li et al. 2019; Tan et al. 2019; Vu et al. 2019). In this paper, we use $N = 63$, that is, only 63 days (3 business months) before the current time point are used for training the Bayesian and the ARIMA models. An illustration of rolling forward method with $N = 26$ is presented in Fig. 4.

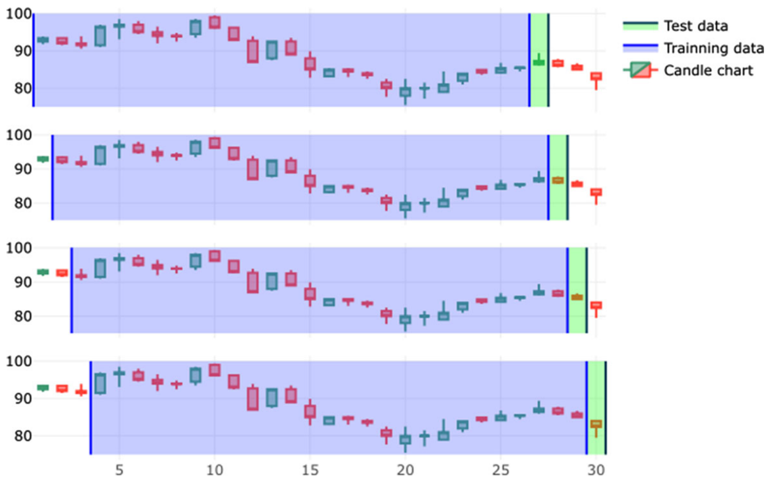


Fig. 4 An illustration for rolling forward method

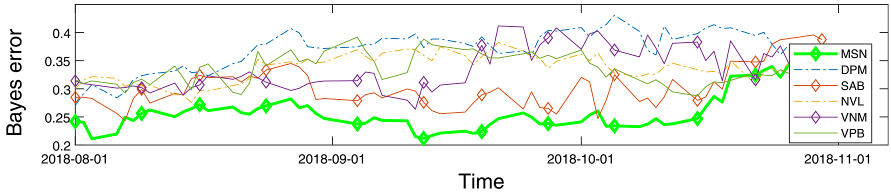
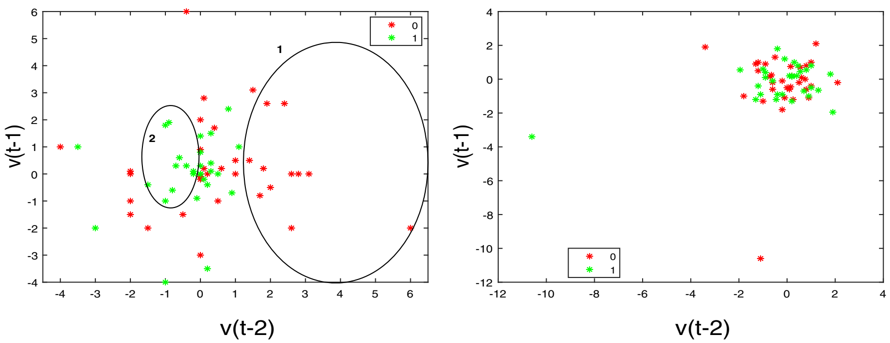


Fig. 5 Bayes error of the six stocks that have the lowest Bayes error

Example 1 This example illustrates the stock selection method, which has been presented in Sect. 3.3. Stocks of VN30 basket collected from May 2, 2018 to October 31, 2018. After transforming the time series data to tabular data using Algorithm 1, we calculate the Bayes error relevant to them. Note that Bayes error can be theoretically estimated without having to make any computation regarding the test data. This error is based on the overlap degree of two density functions. The stocks with lower Bayes error are expected to provide higher probabilities of correct predictions. Additionally, as mentioned earlier, we incorporate the rolling forward method into the training process. Therefore the training data set as well as the Bayes error will change time by time. Figure 5 illustrates the obtained Bayes errors for the six stocks that have the lower Bayes error than the other 24 stocks.

It can be seen from Fig. 5 that the MSN stock has the lowest Bayes error, followed by the SAB stock. Therefore, MSN is the most suitable stock to add in the portfolio by using the proposed method. Figures 6 and 7 illustrate the scatter plots and the estimated probability density functions of MSN and FPT stocks. MSN is the stock that has the lowest Bayes error whereas FPT is one of the stocks having high Bayes error. In Fig. 6, the *x*-axis and *y*-axis represent the variations of the previous two days and one day, respectively. For FPT, it can be seen that the two classes are completely overlapped, no specific rule can be stated, and very high Bayes error is introduced. Therefore, FPT is not suitable for the proposed application. For MSN, it



(a) MSN **(b)** FPT

Fig. 6 Scatter plot of two stocks with legend colored by trend

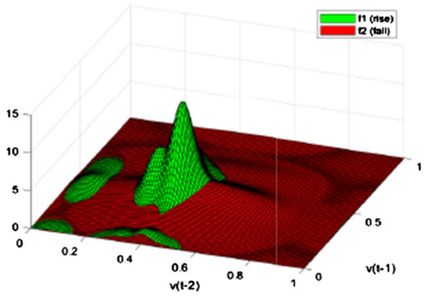


Fig. 7 Estimated probability density functions

can be seen from Fig. 6a that all points in Region 1 are labeled as "fall". Those points have large positive x-values, that is, if stock price sharply rose in the two previous days, it would be likely to fall in current day. Meanwhile, the points in Region 2 are labeled as "rise". The x-values of such points are negative, but near 0, that is, if stock price slightly fell in the previous two days, it would be likely to rise in the current day. Thus, in the case of MSN, the price variation of the previous two days is really useful for predicting the stock prices. The performance can be further improved if both the variations of one and two previous days are used.

In summary, it can be emphasized that all models including the proposed method are only useful in some cases, and their performance mainly depends on the skill of investors, such as identifying the suitable stock and the suitable time for trading, etc. In case of proposed framework, MSN seems to be the most suitable stock. This selection will be evaluated through examples.

Example 2 In this section, the proposed method is applied to predicting stock trend. The prices of MSN stock are collected from May 2, 2018 to October 31, 2018 and the stock prices of the last three months are used as the test set. The performance of proposed method over the test set is compared to three other techniques in terms of profit after transaction cost. The main sources of transaction cost are the commissions and the bid-ask spreads (Kociński 2017; Verousis 2013). In case of Vietnamese stock market, there is no bid-ask spread. Therefore, a commission of 0.15% is used as the transaction cost in this paper. For each technique, we use a fixed order of 10 units once a buy/sell signal occurs. The comparative methods are presented as follows.

- Proposed method: the strategy of decision-making which has been presented in Sect. 3.2.3. The rolling forward method is also incorporated to further improve the predicted performance.
- No-skill algorithm: buy and sell signals are randomly generated.
- Buy and hold strategy: buy the stock on the first day and sell it on the last day of trading period.
- ARIMA model: buy and sell signals are chosen according to the local minimum and maximum prices predicted by ARIMA(p, d, q) model in which p, d, q are

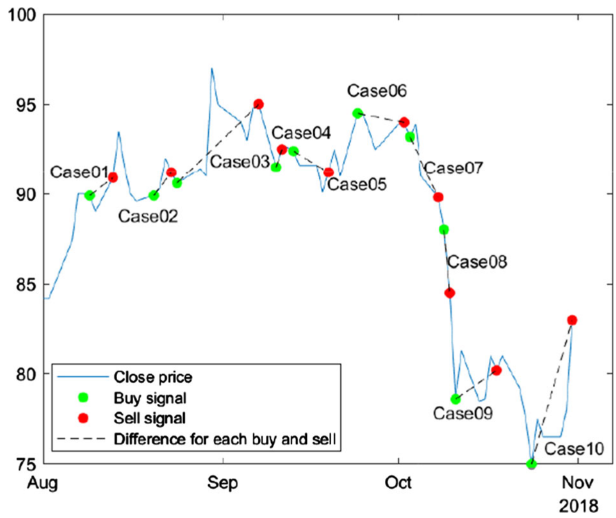


Fig. 8 The stock price, the buy, and sell signal on the test set using the proposed strategy

chosen based on AIC criterion (refer Box and Jenkins 1976 for more details). The rolling forward method is also incorporated into the training process.

- Ichimoku cloud: in this strategy, the cloud is considered as same as the "average region". Thus, a buy signal is identified when the Chikou line crosses the bottom of the cloud from above and a sell signal is identified when the Chikou line crosses the top of the cloud from below (refer Deng et al. 2020; Elliott 2007; Patel 2010; Shawn et al. 2015 for more details).

The obtained results for the comparative methods are illustrated in Figs. 8, 9 and 10. Their profit after transaction cost are summarized in Table 3.

Based on the above results, we have the following remarks.

- It can be observed from Table 3 that the no-skill algorithm achieves a negative profit and a lower performance in comparison with the strategies using Bayesian approach and the Ichimoku cloud. This result evidences that although the stock market is volatile, it still has its own rules; therefore, a random investment strategy is not a good strategy for trading. Even in case that the stock price is a random walk, traders who randomly buy and sell stocks are likely to obtain a negative profit due to the transaction cost.
- The Buy and hold strategy is one of the well-known methods for long-term investigation. It is considered as a smart strategy for investors, especially in the second half of the 20th century. However, recently, there have been many evidences showing that this strategy is no longer suitable due to the increasing volatility of the global financial market and frequent occurring of financial crises (Boscaljon et al. 2008; Hilliard and Hilliard 2015, 2018; Hui and Chan 2019; Ng'ang'a 2019; Sanderson and Lumpkin-Sowers 2018). The obtained result shown in Table 3 reinforces the above viewpoint when this strategy results in a

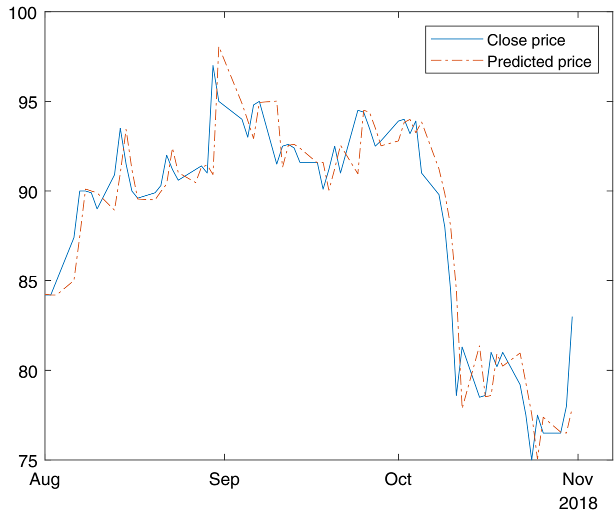


Fig. 9 The stock price and the predicted value of ARIMA

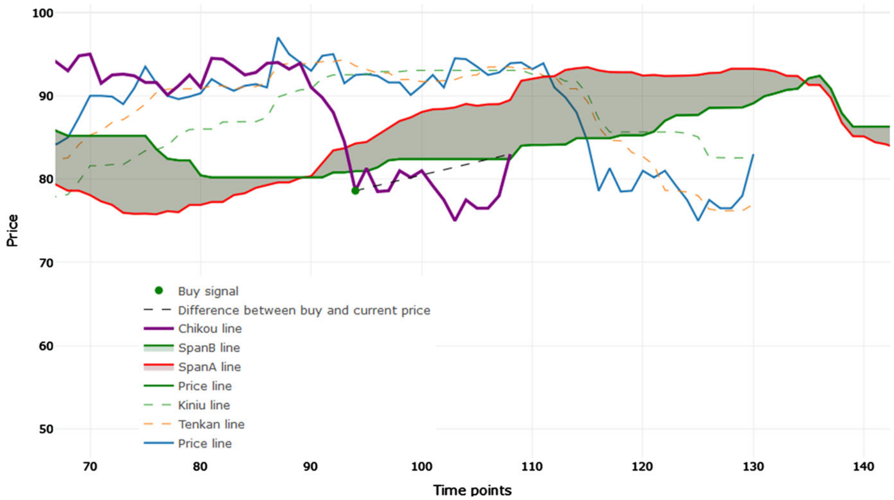


Fig. 10 Ichimoku cloud chart

Table 3 Profits obtained using all methods

	Profit (%)	Note
Proposed method	7.17	
No-skill algorithm	−4.55	
Buy and hold strategy	−1.73	
ARIMA model	−7.72	
Ichimoku cloud	4.94	Has not taken profit already

negative profit. However, it should be noted that the Buy and hold strategy requires a longer investment horizon. Therefore, an investigation over a period of three months might not lead to a reliable conclusion.

- The buy and sell signals using the proposed strategy are shown in Fig. 8. Accordingly, we obtain the positive profit in 6/10 cases. In those six cases, the proposed strategy can well identify the buy and sell signals located near the local minimum and maximum points. Especially, in Case 10, the global minimum price of the trading period is also detected. Nevertheless, the proposed strategy produces incorrect signals in 4/10 cases. Those four cases lie in the range where the stock price suddenly sharply falls. This sudden price movement has not occurred in the earlier training period. As a result, the proposed strategy results in some incorrect predictions within that range, before updating the new training data. This is a disadvantage of the proposed approach. After three months of investigation, the final profit obtained by the proposed strategy is 7.17%, as shown in Table 3. This profit is not too impressive but can be acceptable.
- As shown in Fig. 9, the ARIMA model incorporating the rolling forward method attempts to predict the future stock price movement using some last stock price movements. However, it is likely to be one-step late compared to the actual price. Consequently, the ARIMA model achieves the lowest performance among the comparative methods, as shown in Table 3. This result evidences that stock price is volatile and is not easy to be predicted by a linear model.
- In Fig. 10, the Chikou line crosses the bottom of the cloud from above at time point 94. The Chikou line is as same as the lagging price line, therefore the corresponding time point and price of buying stocks are 116 and 78.6. As shown in Table 3, investors can take a profit of 4.94% until the last day of the test period; however, potential risk can be identified due to the fact that sell signal has not occurred yet and the stock has been held day-by-day.

In summary, it can be observed from Table 3 that the proposed strategy can result in the highest profit among the comparative methods. Based on this result, the proposed method can be considered as a potential method for short-term stock trend prediction and can be considered as a new reference source for investors. The performance of this method also depends on many other factors such as market characteristics, investor's personality traits, and the selected stocks.

Example 3 In this example, the effectiveness of the stock selection method based on Bayes error, which has been presented in Sect. 3.3, will be tested. As implemented in Example 1, MSN stock has lowest Bayes error and is evaluated as the most suitable stock for the proposed strategy. In this example, the proposed strategy is applied to MSN and other stocks in VN30 basket. Similar to Example 2, stock prices are collected from May 2, 2018 to October 31, 2018 and the stock prices in the last three months are used as the test set. Also, we use a fixed order of 10 units once a buy/sell signal occurs. The effectiveness results in terms of one-step ahead prediction error and profit are shown in Table 4.

Table 4 The effectiveness results of the proposed strategy on different stocks

Stock	Error (%)	Profit (%)
MSN	39.06	7.17
SAB	45.31	4.32
VJC	50.00	−1.94
GAS	53.13	−19.22
SSI	46.88	−3.98
VCB	48.48	−2.67
MBB	46.88	−5.85

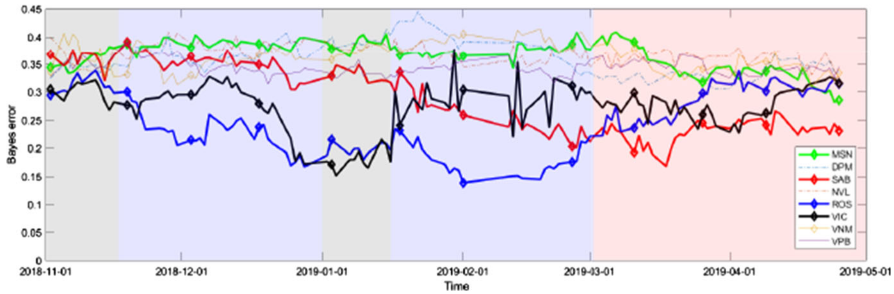


Fig. 11 Bayes error of some stocks in the five test periods

As presented in Example 1, MSN and SAB stocks have the lowest and the second lowest Bayes error, respectively. As shown in Table 4, among the comparative stocks, only these two stocks can achieve positive profit. This demonstrates the suitability of using Bayes error for stock selection. In other words, transforming time series data to tabular data, and predicting stock trend using Bayesian classifier cannot be applied to any stocks, but can be useful for some specific cases. Therefore, selecting a suitable stock using Bayes error plays an important role in the proposed method. Besides, due to the complexity of stock market data, it can be seen from Table 4 that the actual errors have higher values compared to the estimated Bayes errors in Example 1. Therefore, it would be better if investors frequently update the Bayes error during the trading period, follow the Bayes error trend so that they can decide which is the most suitable time and which is the most suitable stock when using this technique. Also, combining the proposed technique with other technical techniques is recommended.

Example 4 In the examples above, we have illustrated the proposed method and presented its performance using stock data from May 1 to October 31, 2018. In this example, the performance of the proposed method is further investigated on the next testing period, namely November 1, 2018 to April 30, 2019. As shown in Fig. 11, the new testing period can be divided into five sub-periods, each with a corresponding stock with the minimum Bayes error. Specifically,

Table 5 Profits and ranking of comparative methods over several periods

	Period 1	Period 2	Period 3	Period 4	Period 5	Ranking average
Proposed	−1.44% (2)	7.62% (3)	0.00% (2)	3.23% (2)	2.64% (3)	2.40
No-skill	−4.56% (4)	9.80% (2)	−1.60% (5)	−2.44% (4)	4.32% (1)	3.20
Buy and hold	−4.27% (3)	5.43% (4)	0.79% (1)	−2.96% (5)	−1.46% (4)	3.40
ARIMA	−7.16% (5)	12.69% (1)	−0.70% (4)	5.28% (1)	−8.32% (5)	3.20
Ichimoku cloud	0.00% (1)	5.10% (5)	0.00% (2)	0.00% (3)	3.10% (2)	2.60

- In Period 1 (from November 1 to November 15, 2018), we invest in VIC stock.
- In Period 2 (from November 16 to December 31, 2018), we invest in ROS stock.
- In Period 3 (from January 1 to January 15, 2019), we invest in VIC stock.
- In Period 4 (from January 16 to February 28, 2019), we invest in ROS stock.
- In Period 5 (from March 1 to April 30, 2019), we invest in SAB stock.

The performance of all comparative methods through all testing periods are summarized in Table 5, where the first number is profit percentage and the number in parentheses is the method ranking in descending order of profit. It can be seen from Table 5 that although the ARIMA, the Buy and hold, and the no-skill method can achieve the highest profit and highest ranking for some specific periods, they rank low for most periods. A common feature of these methods is their instability, that is, each method is only suitable for a specific movement of market. In our point of view, the no-skill algorithm is actually should not be used; the buy and hold strategy is best suited for a market with low prices at the beginning and high prices at the end of the investment period; the ARIMA model is suitable for a market with continuous increases and decreases. The Ichimoku cloud and the proposed method are more stable strategies. The Ichimoku method can achieve non-negative profit at all investment periods. With the lowest number of buy and sell signals, this could be regarded as a conservative investment strategy and better suited for long-term profits. For the proposed strategy, Table 5 indicated that it does not provide the highest profit for all of the five investment periods. However, it is always in the top three strategies with high profitability. As a result, it achieves the best ranking over the five investment periods. This method has a lot of open orders at relatively low prices, and takes profit after a short time when the price is relatively acceptable, thereby avoiding unfavorable movements of the market. In our point of view, this seems to be a conservative investment strategy and better suited for short-term profits.

5 Conclusion

In this paper, a new approach for short-term stock trend prediction is proposed. In particular, time series data are transformed to tabular data and then predicted using Bayesian classifier. In addition to one-step ahead prediction, a method for two-step ahead prediction has been proposed, thereby building a new short-term investment strategy. Finally, a stock selection method based on Bayes error has been introduced. The numerical results have demonstrated the feasibility of the proposed strategy. We also concluded that the proposed strategy could be regarded as a conservative investment strategy and better suited for short-term profits.

References

Addesso, P., Capodici, F., D'Urso, G., Longo, M., Maltese, A., Montone, R., et al. (2013). Enhancing TIR image resolution via bayesian smoothing for IRRISAT irrigation management project. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XV, 8887,* 888710.

Alfarano, S., Lux, T., & Wagner, F. (2005). Estimation of agent-based models: The case of an asymmetric herding model. *Computational Economics, 26*(1), 19–49.

Amin, G. R., & Hajjami, M. (2016). Application of optimistic and pessimistic OWA and DEA methods in stock selection. *International Journal of Intelligent Systems, 31*(12), 1220–1233. https://doi.org/10.1002/int.21824.

Arratia, A. (2014). Time series models in finance. In: *Computational finance. Atlantis studies in computational finance and financial engineering*, vol 1. Atlantis Press, Paris. https://doi.org/10.2991/978-94-6239-070-6_4.

Atsalakis, G. S., Protopapadakis, E. E., & Valavanis, K. P. (2016). Stock trend forecasting in turbulent market periods using neuro-fuzzy systems. *Operational Research, 16*(2), 245–269. https://doi.org/10.1007/s12351-015-0197-6.

Batra, R., & Daudpota, S. M. (2018). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1–5. https://doi.org/10.1109/ICOMET.2018.8346382.

Boscaljon, B., Filbeck, G., & Ho, C. C. (2008). Rebalancing strategies for creating efficient portfolios. *The Journal of Investing, 17*(2), 93–103.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control San Francisco*. California: Holden-Day.

Cartea, A., Jaimungal, S., & Ricci, J. (2014). Buy low, sell high: A high frequency trading perspective. *SIAM Journal on Financial Mathematics, 5*(1), 415–444.

Castellaro, M., Rizzo, G., Tonietto, M., Veronese, M., Turkheimer, F. E., Chappell, M. A., & Bertoldo, A. (2017). A variational Bayesian inference method for parametric imaging of PET data. *NeuroImage, 150,* 136–149.

Chen, C., Dongxing, W., Chunyan, H., & Xiaojie, Y. (2014). Exploiting social media for stock market prediction with factorization machine. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2,* pp. 142–149. https://doi.org/10.1109/WI-IAT.2014.91.

Chen, H. (2008). Stock selection using data envelopment analysis. *Industrial Management & Data Systems, 108*(9), 1255–1268. https://doi.org/10.1108/02635570810914928.

Deng, S., Yoshiyama, K., Mitsubuchi, T., & Sakurai, A. (2015). Hybrid method of multiple kernel learning and genetic algorithm for forecasting short-term foreign exchange rates. *Computational Economics, 45*(1), 49–89.

Deng, S., Yu, H., Wei, C., Yang, T., & Tatsuro, S. (2020). The profitability of ichimoku kinkohyo based trading rules in stock markets and fx markets. *International Journal of Finance & Economics*. 1–16. https://doi.org/10.1002/ijfe.2067.

Elliott, N. (2007). *Ichimoku charts: An introduction to Ichimoku Kinko clouds*. London: Harriman House Limited.

Ghasemiyeh, R., Moghdani, R., & Sana, S. S. (2017). A hybrid artificial neural network with metaheuristic algorithms for predicting stock price. *Cybernetics and Systems, 48*(4), 365–392. https://doi.org/10.1080/01969722.2017.1285162.

Goldman, M. B., Sosin, H. B., & Gatto, M. A. N. N. (2018). Path dependent options: Buy at the low, sell at the high. *The Journal of Finance, 34*(5), 1111–1127. https://doi.org/10.1111/j.1540-6261.1979.tb00059.x.

Gupta, S., & Wang, L. P. (2010). Stock forecasting with feedforward neural networks and gradual data sub-sampling. *Australian Journal of Intelligent Information Processing Systems, 11*(4), 14–17.
Hajjami, M., & Amin, G. R. (2018). Modelling stock selection using ordered weighted averaging operator. *International Journal of Intelligent Systems*. https://doi.org/10.1002/int.22029.
Hilliard, J. E., & Hilliard, J. (2015). Evaluating strategies to maximize portfolio performance measures using rebalancing, buy and hold and monetary policy indicators, forthcoming. The review of pacific basin financial markets and policies.
Hilliard, J. E., & Hilliard, J. (2018). Rebalancing versus buy and hold: Theory, simulation and empirical analysis. *Review of Quantitative Finance and Accounting, 50*(1), 1–32.
Hu, Z., Liu, W., Bian, J., Liu, X., & Liu, T. Y. (2018). Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18, pp. 261–269. ACM, New York, NY, USA. https://doi.org/10.1145/3159652.3159690.
Huang, C. Y., Chiou, C. C., Wu, T. H., & Yang, S. C. (2015). An integrated DEA-MODM methodology for portfolio optimization. *Operational Research, 15*(1), 115–136. https://doi.org/10.1007/s12351-014-0164-7.
Huarng, K., & Yu, H. K. (2005). A Type 2 fuzzy time series model for stock index forecasting. *Physica A: Statistical Mechanics and its Applications, 353,* 445–462. https://doi.org/10.1016/j.physa.2004.11.070.
Hui, E. C., & Chan, K. K. K. (2019). Alternative trading strategies to beat buy-and-hold. *Physica A: Statistical Mechanics and its Applications, 534,* 120800.
Jeon, S., Hong, B., & Chang, V. (2018). Pattern graph tracking-based stock price prediction using big data. *Future Generation Computer Systems, 80,* 171–187. https://doi.org/10.1016/j.future.2017.02.010.
Kale, A., Khanvilkar, O., Jivani, H., Kumkar, P., Madan, I., & Sarode, T. (2018). Forecasting Indian stock market using artificial neural networks. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1–5. https://doi.org/10.1109/ICCUBEA.2018.8697724.
Kociński, M. A., et al. (2017). On transaction costs in stock trading. *Metody Ilościowe w Badaniach Ekonomicznych, 18*(1), 58–67.
Kohli, P. P. S., Zargar, S., Arora, S., & Gupta, P. (2019). *Stock prediction using machine learning algorithms BT–applications of artificial intelligence techniques in engineering* (pp. 405–414). Singapore: Springer.
Li, X., Li, Y., Liu, X. Y., & Wang, C. D. (2019). Risk management via anomaly circumvent: Mnemonic deep learning for midterm stock prediction. arXiv:1908.01112.
Liu, Z., & Zhang, T. (2019). A second-order fuzzy time series model for stock price analysis. *Journal of Applied Statistics, 46*(14), 2514–2526. https://doi.org/10.1080/02664763.2019.1601163.
Maciel, L., & Ballini, R. (2020). Functional fuzzy rule-based modeling for interval-valued data: An empirical application for exchange rates forecasting. *Computational Economics*, *57*, 743–771. https://doi.org/10.1007/s10614-020-09978-0.
Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77–91.
Mladjenovic, P. (2016). *Stock investing for dummies*. Hoboken: Wiley.
Nguyen-Trang, T., & Vo-Van, T. (2017). A new approach for determining the prior probabilities in the classification problem by Bayesian method. *Advances in Data Analysis and Classification, 11*(3), 629–643.
Ng'ang'a, J. (2019). An assessment of select market timing strategies' performance in nairobi securities exchange. Ph.D. thesis, Strathmore University.
Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018). Stock market prediction using machine learning. In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, pp. 574–576. https://doi.org/10.1109/ICSCCC.2018.8703332.
Pätäri, E., Karell, V., Luukka, P., & Yeomans, J. S. (2018). Comparison of the multicriteria decision-making methods for equity portfolio selection: The U.S. evidence. European Journal of Operational Research, 265(2), 655–672. 10.1016/j.ejor.2017.08.001
Patel, M. (2010). *Trading with Ichimoku clouds: The essential guide to Ichimoku Kinko Hyo technical analysis* (Vol. 473). Hoboken: Wiley.
Pham-Gia, T., Turkkan, N., & Vovan, T. (2008). Statistical discrimination analysis using the maximum function. *Communications in Statistics-Simulation and Computation ®*, *37*(2), 320–336.

Pizzo, A., Teyssere, P., & Vu-Hoang, L. (2018). Boosted Gaussian Bayes Classifier and its application in bank credit scoring. *Journal of Advanced Engineering and Computation, 2*(2), 131–138.

Quah, T. S. (2008). DJIA stock selection assisted by neural network. *Expert Systems with Applications, 35*(1), 50–58. https://doi.org/10.1016/j.eswa.2007.06.039.

Roscoe, P., & Howorth, C. (2009). Identification through technical analysis: A study of charting and UK non-professional investors. *Accounting, Organizations and Society, 34*(2), 206–221. https://doi.org/10.1016/j.aos.2008.05.003.

Sanderson, R., & Lumpkin-Sowers, N. L. (2018). Buy and hold in the new age of stock market volatility: A story about etfs. *International Journal of Financial Studies, 6*(3), 79.

Sang, X., Zhou, Y., & Yu, X. (2019). An uncertain possibility-probability information fusion method under interval type-2 fuzzy environment and its application in stock selection. *Information Sciences, 504,* 546–560. https://doi.org/10.1016/j.ins.2019.07.032.

Shawn, K., Yanyali, S., & Savidge, J. (2015). Do ichimoku cloud charts work and do they work better in Japan. *International Federation of Technical Analysts Journal*. 18–24.

Sollis, R., Newbold, P., & Leybourne, S. J. (2000). Stochastic unit roots modelling of stock price indices. *Applied Financial Economics, 10*(3), 311–315. https://doi.org/10.1080/096031000331716.

Syriopoulos, T., Tsatsaronis, M., & Karamanos, I. (2020). Support vector machine algorithms: An application to ship price forecasting. *Computational Economics*, *57*, 55–87. https://doi.org/10.1007/s10614-020-10032-2.

Tan, Z., Yan, Z., & Zhu, G. (2019). Stock selection with random forest: An exploitation of excess return in the Chinese stock market. *Heliyon, 5*(8), e02310.

Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016). Stock market prediction using machine learning techniques. In *2016 3rd International Conference on computer and Information Sciences (ICCOINS)*, pp. 322–327. IEEE.

Vasiliou, D., Eriotis, N., & Papathanasiou, S. (2006). How rewarding is technical analysis? Evidence from Athens stock exchange. *Operational Research, 6*(2), 85–102. https://doi.org/10.1007/BF02941226.

Verousis, T. (2013). Bid-Ask spreads, commissions, and other costs. In H. K. Baker & H. Kiymaz (Eds.), *Market microstructure in emerging and developed markets*. https://doi.org/10.1002/9781118681145.ch18.

Vovan, T. (2017). Classifying by Bayesian Method and Some Applications. In Bayesian Inference, pp. 39–61. InTech.

Vu, H., Van, T. V., Nguyen-Minh, N., & Nguyen-Trang, T. (2019). A technique to predict short-term stock trend using Bayesian classifier. *Asian Journal of Economics and Banking, 3*(2), 70–83.

Wiesinger, J., Sornette, D., & Satinover, J. (2013). Reverse engineering financial markets with majority and minority games using genetic algorithms. *Computational Economics, 41*(4), 475–492.

Xu, Y., & Cohen, S. B. (2018). Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1970–1979.

Yang, F., Chen, Z., Li, J., & Tang, L. (2019). A novel hybrid stock selection method with stock prediction. *Applied Soft Computing, 80,* 820–831. https://doi.org/10.1016/j.asoc.2019.03.028.

Zervos, M., Johnson, T. C., & Alazemi, F. (2012). Buy-low and sell-high investment strategies. *Mathematical Finance, 23*(3), 560–578. https://doi.org/10.1111/j.1467-9965.2011.00508.x.

Zhai, J., & Bai, M. (2018). Mean-risk model for uncertain portfolio selection with background risk. *Journal of Computational and Applied Mathematics, 330,* 59–69. https://doi.org/10.1016/j.cam.2017.07.038.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50,* 159–175.

Zhang, X., & Tan, Y. (2018). Deep stock ranker: A LSTM neural network model for stock selection. In *International Conference on Data Mining and Big Data*, pp. 614–623. Springer.

Zhou, Z., Jin, Q., Xiao, H., Wu, Q., & Liu, W. (2018). Estimation of cardinality constrained portfolio efficiency via segmented DEA. *Omega, 76,* 28–37. https://doi.org/10.1016/j.omega.2017.03.006.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.