

093\_meilany\_UPINCASE\_Ida.ed  
ited.doc  
*by*

---

**Submission date:** 25-Oct-2020 10:18PM (UTC+0700)

**Submission ID:** 1425781794

**File name:** 093\_meilany\_UPINCASE\_Ida.edited.doc (312.5K)

**Word count:** 2238

**Character count:** 11668

# Prediction of Banking Stock Prices Using Naïve Bayes Method

Ida Setiani<sup>1</sup>, Meilany Nonsi Tentua<sup>1</sup>, Sunggito Oyama<sup>1</sup>

<sup>1</sup>Universitas PGRI Yogyakarta, Yogyakarta, Indonesia

[meilany@upy.ac.id](mailto:meilany@upy.ac.id)

**Abstract.** Bank Rakyat Indonesia (BRI) is one of the largest state-owned banks in Indonesia. It has prompted both local and foreign investors to buy shares in the largest bank in the country. Reading stock price trends is very important for investors to buy or sell the shares they own. The method usually used by investors is fundamental analysis and technical analysis. Analysis by reading financial reports is quite complicated and requires high accuracy. Besides that, it takes much time because of the large amount of data available. A banking stock price prediction can be a solution to make it easier for investors to read stock price movements. In this study, the authors will design a web-based system to predict banking stock prices using the Naïve Bayes method. This system can provide investors with ease and effectiveness in reading stock price movements.

**Keywords:** *stock prediction system, naïve Bayes method, web-based*

## 1. Introduction

The stock is instruments of proof of ownership or participation of individuals or institutions in a company [1]. Stocks are volatile; they can go up or down. Shares that have risen and falling prices are also known as fluctuating stocks. Facing stock fluctuations or the ups and downs of stock prices has become commonplace as an investor. Stock fluctuations are normal, but if we overdo it, it can also be a disaster. If a company's stock suddenly drops very low, it can cause losses to many parties [2].

Bank Rakyat Indonesia (BRI) is one of the largest state-owned banks in Indonesia [3]. PT Bank Rakyat Indonesia's Tbk (BBRI) share price on daily stock trading has increased, even approaching the highest price level [4]. Based on trading data from the Indonesia Stock Exchange, BRI's share price in December 2019 rose 1.15% to a level of Rp.4,400 / unit. Stock trading volume reached 833.52 million shares valued at IDR 3.59 trillion. It has prompted local and foreign investors to buy shares in the country's largest bank in the country [5].

Reading stock trends is very important for investors to know where the money is going, whether it becomes more or tends to be less. Investors need to know how to analyze future stock price movements[6]. The method usually used by investors is fundamental analysis and technical analysis. This fundamental analysis uses financial statement data by looking at various financial statements[7]. At the same time, technical analysis can decide by reading a stock price chart in the form of lines. Analysis by reading financial reports is quite complicated and requires high accuracy[8]. Besides that, it takes much time because of the large amount of data available.

Current technology makes it possible to process data precisely, quickly, and accurately. With the help of information technology, all data processing in the system can be done in various places to be

more effective and efficient. The application of techniques in large-scale data processing presents new challenges and techniques which eventually give rise to a new methodology called machine learning[9].

Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases [10]. Machine learning often used for classification include the Naïve Bayes method [11] [12], C45 [13] dan Support Vector Machine [14]. The naïve Bayes method was introduced by a British scientist named Thomas Bayes, where this method successfully answered problems in the fields of probability and statistics [15]. The advantage of using the Naïve Bayes method is that it can self-correct, which means that when data changes, so do changes in the results.

Based on this background, we propose "Prediction of Banking Stock Prices Using the Naïve Bayes Method." It hoped that this system could be used as a medium to facilitate users, especially stock investors, to find out predictions of the movement (up / down) of stock prices. The variables used to determine the classification of the ups and downs of stocks in the study include (1) Lag1 (percentage of return on the previous day), (2) Lag2 (percentage of return on the previous two days), Lag3 (percentage of return for the previous three days), Lag4 (percentage of return for the previous four days), Lag5 (previous five days return percentage), Volume, Today (present return percentage) and Direction (ascending and descending classification).

## 2 Methodology

Naïve Bayes is a simple probabilistic classification that calculates a set of probabilities by summing the frequencies and value combinations from a given dataset. The Bayes theorem will assume all attributes are independent or not interdependent given by values on class variables [15].

Naïve Bayes is based on the simplifying assumption that attribute values are conditionally independent if an output value is given. In other words, given the output value, the probability of observing collectively is the product of individual probabilities. The equation of the Bayes theorem is:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Where :

- X : Data with an unknown class
- H : The data hypothesis is a specific class.
- P(H|X) : The probability of hypothesis H under condition X (posterior probability)
- P(H) : Probability hypothesis H (prior probability)
- P(X|H) : Probability X based on the conditions in the hypothesis H
- P(X) : Probability X

If there is numeric data, then first look for the mean and standard deviation of each parameter that describes the numerical data. The formula used to calculate the calculated average value (mean) can be seen as follows:

$$\mu = \sum_{i=1}^n x_i \quad (2)$$

Where:

- $\mu$  : Data with an unknown class
- $x_i$  : The data hypothesis is a specific class.
- n : The probability of hypothesis H under condition X (posterior probability)

And the equation for calculating the standard deviation (standard deviation) can be seen below:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (3)$$

Dimana :

$\sigma$  : standard deviation

$x_i$  : x value

$\mu$  : average count

3  $n$  : number of samples

To explain the Naïve Bayes theorem, the classification process requires several clues to determine what class is suitable for the sample being analyzed. The following is an example of a Stock Prediction Datamining Calculation Using the Naïve Bayes Algorithm [16].

Calculate the value of the gaussian distribution or normal distribution using the formula:

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma^2 ij}}$$

Where:

$\pi$ : a constant with the value 3.14159. . .

e: exponential number with a value of 2.7183. . .

$\mu$ : average (mean) of the data

$\sigma$ : standard deviation of normally distributed data.

### 2.1. Dataset

The data used in this study are historical share price data at BRI Bank downloaded from yahoo.finance for the last five years. 1300 data have been downloaded. The data is divided into two parts, namely, 70% of the data is used as train data and the rest is used as testing data.

### 2.2. Data Flow Diagram

DFD level 1 consists of three main processes, namely admin login, master data, and Naïve Bayes analysis. The system carries out the admin login process by displaying the login session and carried out by the admin by entering a username and password. The admin is carried out master data process by inputting the stock price data then. The admin gets the output from the system in the form of stock price data and stock price predictions. The Naïve Bayes analysis process accepts input in the form of adj comparative data, which is processed from the admin and the user. The output of this process is the result of a Naïve Bayes analysis. DFD level 1 on the system can be seen in Figure 1.

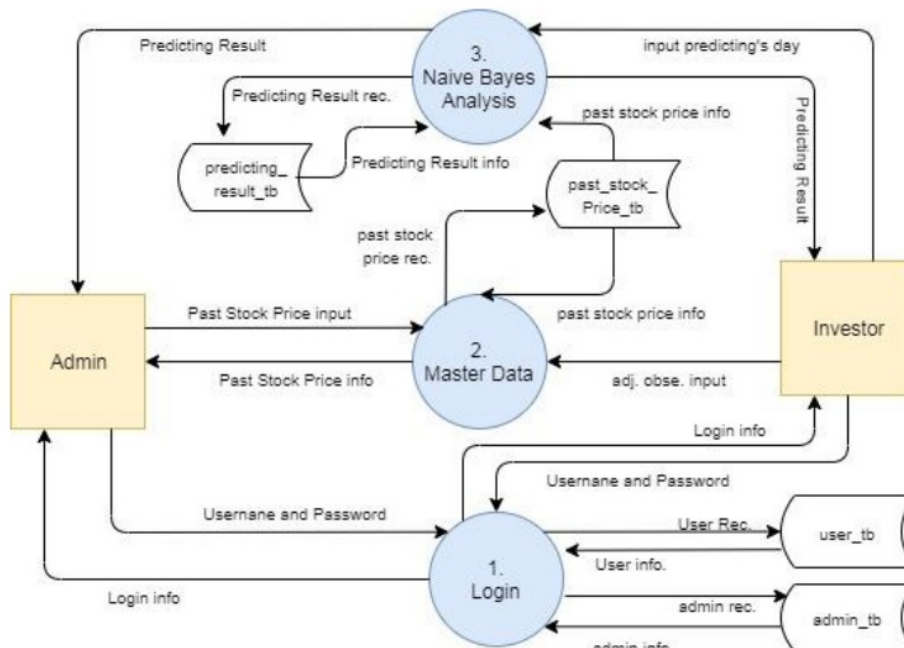


Figure 1. DFD First Level

### 3. Result and Discussion

The prediction system for banking stock prices using the naïve Bayes method can provide results in predicting stock price classifications based on previous stock data. This system can be a reference for users and investors in knowing the movement (up / down) of stock prices. This system consists of several pages, the login page, past share page, prediction page.

The admin login page is the page used by the admin to perform the admin login process. The admin login form is used as protection to protect the admin form from unauthorized people. This form functions to enter a username and password. Then the system will check whether the username and password entered matches the database. The last page is a page used to upload stock price data, which is used as train data.

The user's prediction results page is a page that is used to display the output in the form of calculation results using the Naïve Bayes method. The results of the calculation of the stock price prediction are displayed in table form. These results include predicting stock prices on the next day (H + 1) to the next ten days (H + 10). The predicting page can be seen in figure 2.

No	Tanggal	Hasil Prediksi	Klasifikasi
1	29-08-2020	0.391890326999504	Naik
2	13-08-2020	0.47720948618132	Turun
3	14-08-2020	0.84822101643832	Naik
4	15-08-2020	0.411989797131716	Turun
5	16-08-2020	0.92402475822726	Turun
6	17-08-2020	0.87487269630976	Naik
7	18-08-2020	0.70285433056881	Naik
8	19-08-2020	0.742908394028842	Naik
9	20-08-2020	0.785819870906861	Naik
10	21-08-2020	0.821209022971188	Naik

Figure 2. Predicting Page

The evaluation test content several questions, which are : the ease of using the program and the suitability of input with the information generated (output). Based on the results of the questionnaire about the appearance of the program, the following information was obtained, 27% of respondents answered very interestingly, 60% of respondents answered interesting, and 13% answered less attractive. The evaluation diagram can be seen in Figure 3.

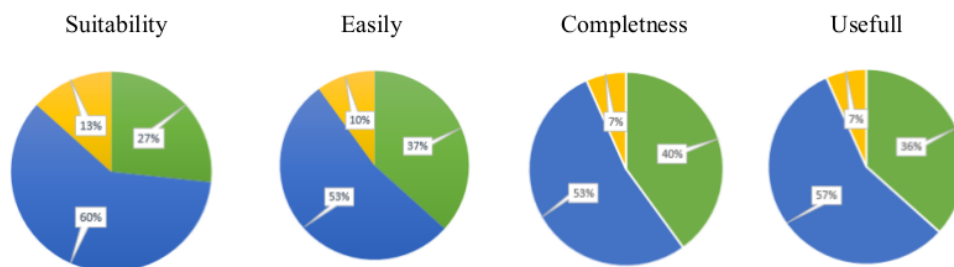


Figure 3. Evaluation System

Based on the questionnaire about the ease of using the application, the following information was obtained, 37% of respondents answered very easily, 53% respondents answered easily, 10% of respondents answered less easily. Based on the results of the questionnaire about the completeness of the information presented, the following data were obtained, 40% respondents answered very well, 53% respondents answered well, 7% respondents answered poorly. Based on the questionnaire about the benefits of the application for users, the following data were obtained, 36% of respondents answered very usefully, 57% useful. Accuracy on the model is calculated by matching the system's predicted results with the actual results on the testing data. With the model obtained in the system, the model's accuracy rate of the model is obtained at 57%.

#### 4. Conclusions

Prediction of Banking Stock Prices Using the Naïve Bayes Method has produced a banking stock price prediction system to assist in considering decision making for stock investors. Prediction results can

be used as a medium to facilitate users, especially stock investors, to find out predictions of stock price movements (up / down).

Testing the application of the banking stock price prediction system using the Naïve Bayes method show that the system can run well. It can be proven by the highest value of the results of the trials carried out, namely the application display trial, 60% of respondents answered interestingly, the ease of running the program 53% of respondents answered very easily, 53% of respondents answered program performance well, and 57% of respondents answered usefully. Accuracy on the model is calculated by matching the system's predicted with the actual results on the testing data. With the model obtained in the system the accuracy rate of the model is obtained at 57%.

### References

- [1] F. Ruhani, T. Salha, and T. Ahmad, "Theories Explaining Stock Price Behavior : A Review of the Literature," *Int. J. Islam. Bank. Financ.*, vol. 2, no. February 2019, 2018.
- [2] S. A. Setiawan, "Does Macroeconomic Condition Matter for Stock Market ? Evidence of Indonesia Stock Market Performance for 21 Years," *Indones. J. Dev. Plan.*, vol. IV, no. 1, pp. 27–39, 2020.
- [3] "Digital Banking - Bank BRI | Melayani Dengan Setulus Hati." [Online]. Available: <https://bri.co.id/digital-banking>. [Accessed: 15-Oct-2020].
- [4] A. Rizkiana *et al.*, "Cogent Economics & Finance The development of composite sentiment index in Indonesia based on the internet-available data The development of composite sentiment index in Indonesia based on the internet-available data," *Cogent Econ. Financ.*, vol. 7, no. 1, 2019.
- [5] D. Hoang, B. Phan, T. Thao, N. Nguyen, and D. T. Nguyen, "A Study Of Indonesia ' S Stock Market : How Predictable Is It ?," *Bull. Monet. Econ. Bank.*, vol. 21, pp. 465–476, 2019.
- [6] M. Asaduzzaman, "New Decision Making Algorithms for Stock Market," University of Fukui, 2015.
- [7] A. Atkins, M. Niranjan, and E. Gerding, "Financial news predicts stock market volatility better than close price," *J. Financ. Data Sci.*, vol. 4, no. 2, pp. 120–137, 2018.
- [8] J. Xianya, H. Mo, and L. Haifeng, "Stock Classification Prediction Based on Spark," *Procedia Comput. Sci.*, vol. 162, no. Itqm, pp. 243–250, 2019.
- [9] N. Milosevic, "Equity forecast : Predicting long term stock price movement using machine learning," Manchester.
- [10] R. Turban, E., Rainer, R. and Potter, *Introduction to Information Technology gy Turban Chapter*. 2003.
- [11] S. Mbadi, "Predicting Stock Market Movement Using Naïve Bayes Model for Sentiment," Jomo Kenyatta University of Agriculture and Technology (JKUAT) Department, 2018.
- [12] C. Huang and Y. Liu, "Machine Learning on Stock Price Movement Forecast : The Sample of the Taiwan Stock Exchange," *Int. J. Econ. Financ. Issues*, vol. 9, no. 2, pp. 189–201, 2019.
- [13] M. N. Tentua and A. Sihabuddin, "Improved C4 . 5 Algorithm Using The L ' hospital Rule And Prunning On The Recommendation System," *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, 2019.
- [14] M. Santoso, R. Sutjiadi, and R. Lim, "Indonesian Stock Prediction using Support Vector Machine ( SVM )," in *MATEC Web of Conferences*, 2018, vol. 01031, pp. 5–9.
- [15] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*, vol. 9781107057. 2013.
- [16] R. T. Gareth James, Daniela Witten, Trevor Hastie, *An Introduction to Statistical Learning with Applications in R*. New York: Springer Texts in Statistic, 2013.



ORIGINALITY REPORT

---

13%

SIMILARITY INDEX

8%

INTERNET SOURCES

13%

PUBLICATIONS

11%

STUDENT PAPERS

---

PRIMARY SOURCES

---

- 1** Mustakim, Reysa Hastarimasuci, Petir Papilo, Zarkasih, Zaitun, Alwis Nazir. "Variable Selection to Determine Majors of Student using K-Nearest Neighbor and Naïve Bayes Classifier Algorithm", Journal of Physics: Conference Series, 2019 2%  
Publication

---
- 2** N L W S R Ginantra, I G A D Indradewi, E Hartono. "Machine learning approach for Acute Respiratory Infections (ISPA) prediction: Case study Indonesia", Journal of Physics: Conference Series, 2020 2%  
Publication

---
- 3** Nadia Ardhianie, Rachmadita Andreswari, Muhammad Azani Hs. "SENTIMENT ANALYSIS OF 'INDONESIAN NO DATING CAMPAIGNS' ON TWITTER USING NAÏVE BAYES ALGORITHM", 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), 2019 2%  
Publication

---



4

Submitted to Asia Pacific University College of  
Technology and Innovation (UCTI)

Student Paper

2%

5

[scitepress.org](https://scitepress.org)

Internet Source

2%

6

Fitriana Harahap, Ahir Yugo Nugroho Harahap,  
Evri Ekadiansyah, Rita Novita Sari, Robiatul  
Adawiyah, Charles Bronson Harahap.

"Implementation of Naïve Bayes Classification  
Method for Predicting Purchase", 2018 6th  
International Conference on Cyber and IT  
Service Management (CITSM), 2018

Publication

1%

7

Submitted to Universitas Amikom

Student Paper

1%

8

Meilany Nonsi Tentua, M. Fairuzabadi,  
Saptaningsih Sumarmi. "Mobile Web-Based  
Information System for thesis's Administration in  
University PGRI Yogyakarta", Journal of  
Physics: Conference Series, 2019

Publication

1%

9

[www.poskerja.com](http://www.poskerja.com)

Internet Source

1%

Exclude bibliography  On