**Predicting Stock Price Trend Using Neural Network**

**Based on Financial Lexicon and Technical Indicators**

Liang-Ching Pan

Department of Information Engineering and Computer Science, Feng Chia University

No. 100, Wenhua Rd., Xitun Dist.

Taichung City 407, Taiwan

liangchingpan1022@gmail.com

**Abstract**

This study uses multiple linear regression (MLR) model and artificial neural network (ANN) model to predict stock price trend, taking TSMC, a company with a large amount of news, as the study object. Collecting TSMC historical stock price, news, and financial statements. we crawl financial news published by various newspapers from January 1, 2017 to December 31, 2021. We tag financial news as two major categories: "TSMC Related News" and "Market Related News” and calculate the news sentiment scores of the two major categories using a customized sentiment analysis dictionary.

**Key words:** Sentiment analysis, Artificial neural network, Linear regression, Stock price prediction

**Introduction**

Since investing in stocks can make considerable profit, stock price trend prediction has become a popular academic research topic. A large number of academic papers related to stock price prediction have been published worldwide. Most of the stock trend prediction researches are conducted in machine learning due to the development and prevalence of it. [7] is market-level research: Using news sentiment as a correlation between companies and stock prices to predict stock price trends. [8] is industry-level research: Studying all three industries - IT, banking, and healthcare - MARS has proven to be the best performing model for stock forecasting in the study. [9] is company-level research: Building up neural network models for short-term technical analysis to study TSMC.

Except for the difference between the size of the research subjects, these inputs of stock price trend prediction models also differ from their features, which may be news analysis [10], fundamental analysis, and technical analysis [9], and different data processing will be done to make a dataset.

The development of internet technologies has made it easier for investors to access stock market information through media. The impact of news on the stock market has three main aspects: (1) fundamental information in company-specific news articles affects investors' trading activities; (2) news evokes public sentiment, and investors' decisions are influenced by public sentiment and thus interfered with investment decisions; (3) the impact of online media on stocks varies depending on news content and company (3) the impact of online media on stocks varies with the content of the news and the characteristics of the company [11].

According to [12], fundamental analysis is based on three basic aspects (1) macroeconomic analysis, such as Gross Domestic Product (GDP) and Consumer Price Index (CPI), to analyze the impact of the macroeconomic environment on the company's future profits, (2) industry analysis, to estimate the value of the company based on the current status and future of the industry, and (3) company analysis, to analyze the current operational and financial status of the company to assess its internal value.

There are many fundamental analysis indicators, technical analysis indicators and others like GDP, CPI can be the features of machine learning, thus [13] used decision trees and multiple regression methods to predict the banking industry and the results showed that the reduction of input variables had a positive impact on the predictive performance of the model.

There are many factors that affect stock price movements. In terms of size, the economy, industry, and individual companies all have different influencing factors, and investors need to obtain information about the economy, industry, and individual companies through news and information regularly to make investment decisions.

In addition to news, investors can also judge the current operating conditions of a company from a single company's accounting statements. To make it easier for investors to interpret accounting data, investors will calculate information using EPS, ROE, gross margin, etc. These analytical indicators can be easily obtained from the Internet without investors having to do their own calculations or graphs, so it is a stock price analysis tool often used by many investors.

Historical stock prices like open, min, max, close price etc. can also be used as a reference in short-term investment decisions, investors may use recent stock price movements to determine when to invest.

This study aims to use news information, analytical indicators, and historical stock prices as stock price references, and to use multiple linear regression (MLR) and artificial neural network (ANN) to develop a single-company stock price trend model to support investors to invest in stocks.

We collects historical stock prices, stock market news, and financial statements as datasets, and examines the stock market data of Taiwan from January 1, 2017 to December 31, 2021 to investigate whether the prediction method of stock prices using multivariate feature input models is appropriate; and whether different methods of cutting datasets can improve the prediction accuracy of the models. We also calculate root mean squared error (RMSE), accuracy, precision, recall, and F1-Score to evaluate the model effectiveness of this study, in order to train the most suitable model for predicting stock prices.

**Literature Review**

*A. Multiple Linear Regression, MLR*

A linear model with ordinary least squares linear regression fitted with coefficients , it minimizes the sum of squared error between the observed and predicted targets by linear approximation. It is widely used in the case of machine learning for sequential or categorical models.

In the case where the input variable is single, the linear regression is called simple linear regression; in the case where the inputs are multiple variables, the linear regression is called multiple linear regression. Multiple linear regression finds the sum of squared error through iterating.

is feature, is coefficient, is bias and is predict target.

*B. Artificial neural network, ANN*

The computing power of computers has increased dramatically that computers can afford the huge matrix computing capacity of neural network, resulting in the prevalence of machine learning. Neural networks can be applied to time series prediction and classifiers, which can input multiple features, and the types of data can be text, speech, video, etc. The number of neurons, the number of hidden layers, and the type of activation functions can be changed to achieve better prediction results.

The artificial neural network itself is backward propagation neural network because it updates the model by minimize the loss function by iterations, so that the model coefficients are updated in the direction of minimizing the loss function. It is characterized by fast iteration speed, high learning accuracy, and the ability to handle nonlinear relational data [9].

*C. News Sentiment Analysis*

News sentiment can influence investors and cause market volatility [11]. News articles can reveal individual companies, the overall market situation, and have both short-term and long-term effects. By analyzing the news, we can get some useful information that can be used as a factor to influence the stock price.

*D. Dictionary-based Sentiment Analysis*

Dictionary-based sentiment analysis is often used in literature as a feature extraction method. A dictionary contains words with specific features, and the more precise word is, the more precisely specific features in a text can be extracted. Specific features are often classified as positive or negative. For example, in [10], the specific feature of the dictionary is stock price fluctuation, and the positive implication for stock market news is that the stock price is going up; the negative implication is that the stock price is going down. The sentiment score is calculated from the word frequency, we predict stock price will go up if the number of positive sentiment words deduct the number of negative sentiment words > 0, and we predict stock price will go down if the number of positive sentiment words deduct the number of negative sentiment words < 0. Both [10] and [14] used the same approach to extract sentiment features from texts. whie

**Research Method**

Position figures and tables are at the second page, if possible. Large figures and tables may span both columns. Figure captions should be below the figures; table captions should be above the tables. Use the abbreviation (e.g., “Fig. 1”) even at the beginning of a sentence.

All half-tone illustrations (pictures/photographs) should be clear black and white prints. Do not use photocopies. These illustrations should be furnished within the copy. Make certain to include a caption in the paper for the illustration as well as to label the illustration on the back.

**Helpful Hints**

*A. References*

List and number all references at the end of the paper. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [1]. Number the citations consecutively. The sentence punctuation follows the brackets. Do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence.

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it is cited. Do not put footnotes in the reference list.

Give all authors’ names; do not use “et al” unless there are three authors or more. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5].

For papers published in translated journals, please give the English citation first, followed by the original foreign-language citations [6].

*B. Abbreviations and Acronyms*

Define abbreviations and acronyms the first time they are used. Acronyms such as MOSFET, ac and dc do not have to be defined. Redefine acronyms when first used in the text, even if they have been defined in the abstract.

*C. Equations*

Number equations consecutively with equation numbers in parentheses flush with the right margin, as in (1). To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Use parentheses to avoid ambiguities in denominators. Punctuate equations with commas or periods when they are part of a sentence, like this,

  (1)

|  |
| --- |
| this is  a sample  figure |

Fig. 1 This is a sample figure. Captions exceeding one line are arranged like this.

Be sure that the symbols in your equation have been defined before the equation appears or immediately following. When

TABLE I

FONT SIZES FOR CAMERA-READY PAPERS

|  |  |  |  |
| --- | --- | --- | --- |
| Font Size | Bold | Italic | Text |
| 10 |  |  | Main text, authors’ affiliations |
| 10 | Yes |  | Headings, i.e. **Abstract** |
| 12 |  |  | Authors’ names |
| 14 | Yes |  | Paper title |
| 10 |  | Yes | Sub-headings, i.e., *Fonts* |
| 9 |  |  | References, table, table names, table captions, figure captions |
| 8 |  |  | Footnotes, sub- and superscripts |

you refer to equations in the text, refer to (1). Do not use “Eq. (1)” or “Equation (1)” except at the beginning of a sentence: “Equation (1) is used....”

*D. Other Recommendations*

Use one space between sections, if possible.

**References**

[1] G. Eason, et al., *Phil. Trans. Roy. Soc. London,* A247, pp. 529-551, April 1955.

1. J. Clerk Maxwell, *A Treaties on Electricity and Magnetism*, 3rd ed., Vol. 2, Oxford: Clarendon Press, 1892, pp. 68-73.
2. I. S. Jacobs and C. P. Bean, *Magnetism*, 3, G. T. Rado and H. Suhl, Eds., New York: Academic Press, 1963, pp. 271-350.
3. M. Smith, “Title of paper optional here,” unpublished.
4. K. Rose, “Title of paper with only first word capitalized,” in press.
5. Y. Yorozu, et al., *IEEE Trans. J. Magn. Japan*, 2, pp. 740-741, August 1987 [*Digests 9th Annual conf. Magn. Japan*, p. 3012, 1982]