Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/45/20230275

Stock Price Forecasting with Machine Learning

Shichen Ying[1,a,*]

[1]Wuhan University of Technology, Wuhan Hubei 430070, China
a. 299758@whut.edu.cn
**corresponding author*

Abstract: Stock price prediction has always been a problem that investors are very concerned about. This paper studies the use of financial information of listed companies to predict whether the stock price will rise after the release of its financial report. This paper extracts the financial information of listed companies from the Guotaian database, obtains their stock price data from Yahoo Finance, and calculates the corresponding technical indicators. And exploring the effect of these indicators on the stock price prediction. The study found that there is a small gap between forecasting with financial information alone and forecasting with technical indicators alone. The combined model performs slightly better than the single model. This study demonstrates that financial information can effectively aid in predicting stock prices and overcome the limitations of certain technical indicators. By incorporating both financial data and stock price information, investors can make more accurate prediction regarding fluctuations in the stock market.

Keywords: machine learning, stock price prediction, financial indicators

1. Introduction

The stock market is an extremely important part of a country's economy, and its importance lies not only in meeting the capital needs of investors but also in the important role it plays in economic growth, corporate finance, industrial structure, and social stability. The stock market, as the main channel of corporate financing, helps to increase the capital of enterprises and improve their investment capacity, thus promoting economic growth.

The stock market is a speculative field where high investment returns are often accompanied by high risks, a characteristic that has prompted people to seek ways to achieve high returns while minimizing risks.

Historically, there have been two categories of stock price forecasting, fundamental analysis, and technical analysis. Among them, technical analysis focuses purely on the simplest patterns of changes in supply and demand in the financial markets, constructing indicators from trading data to predict stock price movements, while fundamental analysis explores and evaluates the intrinsic value of stocks [1]. In addition, there are methods to solve the stock price forecasting problem using statistical methods, especially statistical time series models, such as ARIMA, GRACH, and other models, which solve the problem to some extent. For example, Wu and Wen developed an ARIMA model to predict the direction and pattern of changes in the closing price of Huatai Securities [2]; Lin used a GRACH model to model stock volatility [3]; and Zhang incorporated approximate differential information on

DOI: 10.54254/2754-1169/45/20230275

the lagged value of stock prices into the ARMA-GRACH model to improve the prediction accuracy [4].

Overall, these methods address the problem to some extent. Traditional analytical methods consider only part of the information, and the analytical judgment relies on human experience, which has the problem of being too subjective, while statistical methods, which require high constraints to be satisfied by the data, are not too effective for long-term prediction and generally require preprocessing such as legitimization or differencing for nonlinear nonstationary series [5].

Machine learning methods, on the other hand, have unique advantages. Unlike previous methods, machine learning can accept multivariate inputs, which means that it can receive information other than past stock prices. There is a large literature showing the impact of financial indicators on stock prices [6-10]. The models chosen for this paper are logistic regression, support vector machine, random forest, and XGBoost. This paper hopes to investigate whether increasing the use of financial indicators for forecasting will lead to a better model performance by examining the performance of each machine learning model when financial indicators, technical indicators, and both are used.

In this paper, quarterly financial data and stock price data of Chinese listed companies over the past 10 years were obtained from the Guotaian database (CSMAR). First, this paper performs data cleaning by filling in missing values and data normalization. Second, the data is split into a training set and a validation set before constructing the model. Thirdly, the prediction results of each model are compared on the validation set. The performance of each model and different training data is evaluated based on selected indicators. In this paper, the result was found that the use of technical and financial indicators had little effect on the performance of the models, with the models with technical indicators performing slightly better. The performance of the models using both is slightly improved compared to the models using technical and financial indicators alone. This shows that incorporating financial indicators into a machine-learning model for predicting stock prices can enhance the model's performance.

This paper is structured as follows: Section 2 provides a description of the selected model, including its characteristics and functionality. Section 3 discuss the data used in this study, their sources, and how they were processed to create the measures for analysis. In the following section 4, the article presents and describes the results and analyzes the differences between the different methods and models. Finally, the conclusion is made by describing the findings and contributions.

2. Model

In this paper, logistic regression, support vector machine, random forest, and XGBoost are chosen for the models.

2.1. Lgistic Regression

Logistic Regression is a common classification algorithm that classifies the input features x and weights w through a Sigmoid function $\sigma(z)$ that can restrict the value domain between [0, 1], i.e. mapping the linear output to the probability space.

Suppose the input features of the training sample are $X = [\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{N}]^T$, and the corresponding output is $y = [y_1, y_2, \dots, y_N]^T$, where y only takes the values 0 or 1, and x and w are d-dimensional vectors, respectively, then the model of logistic regression can be expressed as:

$$P(y = 1|\mathrm{x}, \mathrm{w}) = \sigma(\mathrm{w}^T\mathrm{x}) = \frac{1}{1 + exp^{-\mathrm{w}^T\mathrm{x}}} \tag{1}$$

where the Sigmoid function $\sigma(z)$ is defined as follows:

DOI: 10.54254/2754-1169/45/20230275

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{2}$$

The loss function of logistic regression is defined as Cross Entropy Loss, which can be expressed as:

$$\mathcal{L}(\mathrm{w}) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\sigma(\mathrm{w}^T \mathrm{x_i})\right) + (1 - y_i)\log\left(1 - \sigma(\mathrm{w}^T \mathrm{x_i})\right)\right] \tag{3}$$

where y_i denotes the category to which sample i belongs, and the log-likelihood is represented by two log functions in Eq.

In order to optimize the parameters w in training, the average cost (loss function) needed to minimize over the training samples. In the optimization process, Stochastic Gradient Descent (SGD) can be used to update one randomly selected sample at a time. The specific update process can be expressed as follows:

$$\mathrm{w} \coloneqq \mathrm{w} - \alpha[(\hat{y} - y)\mathrm{x}] \tag{4}$$

where $\hat{y} = \sigma(\mathrm{w}^T \mathrm{x})$ denotes the predicted value, y denotes the sample true label, and α denotes the learning rate. The above equation can be used to update the weights w.

With the training data and optimization process, decision bounds can be obtained, as well as a logistic regression classifier. For a new input feature sample, the logistic regression classifier predicts the probability of belonging to a category based on the Sigmoid value it obtains.

2.2. Support Vector Machine

Support Vector Machine (SVM) is a model based on statistical learning theory, mainly used for classification and regression analysis.

Suppose there is a set of observed samples with known categories $(\mathrm{x_1}, y_1), (\mathrm{x_2}, y_2), \dots, (\mathrm{x_N}, y_N)$, where each sample $\mathrm{x_i}$ is a d-dimensional vector and the value of y_i is -1 or 1 indicating the category of the corresponding sample. The goal of the SVM is to find a hyperplane that separates these samples into two classes and maximizes the distance (called Margin) to that hyperplane for the samples that support it.

The hyperplane is defined by the equation $\mathrm{w}^T\mathrm{x} + b = 0$, where w represents the normal vector and b denotes the intercept. Considering that the classification problem is a regularized minimization problem, the optimization problem can be expressed as:

$$\min_{\mathrm{w,b,\xi}} \frac{1}{2}\|\mathrm{w}\|^2 + \mathrm{C}\sum_{\mathrm{i=1}}^{\mathrm{N}} \xi_{\mathrm{i}} \tag{5}$$

where $\|\mathrm{w}\|$ is the L2 parametrization of the d-dimensional vector w, $C > 0$ is a pre-specified regularization parameter, and ξ_i is a relaxation variable indicating the degree to which the i-th sample can be misclassified.

In order to constrain the maximum value of Margin, each sample must satisfy the following constraints:

$$\mathrm{y_i}(\mathrm{w^T x_i} + \mathrm{b}) \geq 1 - \xi_{\mathrm{i}}\xi_{\mathrm{i}} \geq 0 \tag{6}$$

DOI: 10.54254/2754-1169/45/20230275

Based on the model and constraints specified above, the Lagrange multiplier method can be used to optimize the objective function of the SVM by solving for the support vectors. After solving, a hyperplane can be obtained that allows the classifier to achieve the best performance while solving for the inter-sample interval maximization.

2.3. Random Forest

Random forest is an integrated learning method that aims to improve the accuracy and stability of classification, regression and other tasks by combining multiple decision trees. The randomization process in it mainly includes sample randomization and feature randomization.

Assuming there are N samples, K features and T trees. The training set is represented by $D = (x_1, y_1), \ldots, (x_N, y_N)$, where x_i denotes the feature vector of the i-th sample and y_i represents its corresponding label.

The training process of a random forest can be divided into two randomization steps:

For each tree t and training set D: Randomly sample the training set D with put-back to obtain a subset D' of size N'. For each sample point i and each feature k, a threshold $\theta_{i,k}$ is randomly generated, and samples smaller than the threshold are assigned to the left subtree and samples larger than or equal to the threshold are assigned to the right subtree to obtain a binary decision tree. The CART decision tree algorithm is used here, but the Gini index is used for the classification problem and the mean square error is used for the regression problem.

For each feature k, a threshold θ_k is randomly chosen, and the k feature of all sample points that is less than the threshold is grouped into the left subtree, and those that are greater than or equal to the threshold are grouped into the right subtree.

Define the attribute set J, where $|J| = m$. Let the Gini index of the current node sample set be G and the Gini indices of the child node sample sets be G_L and G_R, respectively. Each time a node splits, m attributes are randomly selected from the attribute set J, and then the best attribute is selected for splitting, and the splitting is recursive to the left and right child nodes respectively.

The final prediction result can be obtained by combining the prediction results of all decision trees, for example, for classification problems, the category with the most votes can be used as the final prediction result.

2.4. XGBoost

XGBoost is an excellent integrated learning method which is based on decision tree boosting algorithm. XGBoost enhances the predictive performance of the model while mitigating the risk of overfitting through the utilization of gradient boosting and regularization techniques.

Suppose there are N training samples, K features and T decision trees, and the training set is $D = (x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, where x_i is the feature vector of the i-th sample and y_i is the label of the i-th sample.

XGBoost's primary optimization objective is to minimize the objective function, which comprises two components: the loss function and regularization term. The logistic loss function is utilized for classification problems, while the squared loss function is used for regression problems.

The expression of the objective function is given by:

$$
\begin{aligned}
Obj(\theta) &= L(\theta) + \Omega(\theta) \\
&= \sum_{i=1}^{N} l\,(y_i, \hat{y}i) + \sum j = 1^T \Omega(f_j) \\
&= \sum_{i=1}^{N} l\,(y_i, \hat{y}i) + \sum j = 1^T \left[\gamma_T T + \frac{1}{2}\lambda |w|^2\right] \\
&= \sum_{i=1}^{N} l\,(y_i, \hat{y}i) + \sum j = 1^T \left[\gamma_T T + \frac{1}{2}\lambda \sum_{k=1}^{K_j} \omega_{j,k}^2\right]
\end{aligned}
\tag{7}
$$

The equation includes several terms and parameters. $L(\theta)$ represents the loss function, while $\Omega(\theta)$ is the regularization term. $\hat{y}_i$ denotes the predicted value for sample i and $f_j(x_i)$ represents the predicted value for tree j. The loss function for each sample is denoted by $l(y_i, \hat{y}i)$. Additionally, γ_T and λ are parameters related to tree structure and leaf node weight, respectively. The weight vector of the model is represented by w , while $\omega_{j,k}$ refers to the weight of leaf node k node in tree j.

The optimization method of the objective function uses a gradient boosting technique, which means that each iteration calculates a residual based on the error between the predicted and actual values of the current model, and then retrains a new decision tree with the residual as the new target value to be added to the current model. In this way, it is possible to iterate continuously and build multiple decision trees to improve the fitting and generalization performance of the model.

To prevent overfitting, a regularization term is included in the objective function. This limits the complexity of the tree and consists of two parts: the tree structure parameter and leaf node weight parameter. The regularization term aims to improve the model's generalization ability by preventing it from overfitting on training data.

3. Data Source

For this study, financial indicator data was obtained from the Guotaian database by selecting industry code C39. This code pertains to the data of companies listed in the industries of computer, communication, and electronic equipment manufacturing. Stock price data was sourced from Yahoo Finance (finance.yahoo.com). TaLib was used to calculate relevant indicators based on the stock price data obtained.

This study examines the use of machine learning algorithms to predict whether a company's stock price will increase the day after the release of a public company's financial statements.

The technical indicators used in this study are the weighted moving average (WMA), the applied double exponential moving average (DEMA), the average convergence index (ADX), the exponential smoothed dissimilarity moving average (MACD), the homeopathic indicator (CCI), the momentum indicator (MO), and the relative strength indicator (RSI). The above indicators are listed in Table 1.

DOI: 10.54254/2754-1169/45/20230275

Table 1: Introduction of technical indicators.

Indicators	Description
WMA	WMA stands for Weighted Moving Average, which is a popular technical analysis indicator used in financial markets. It is a type of moving average that places greater importance on recent data points while still taking into account past data.
DEMA	DEMA stands for Double Exponential Moving Average, which is a technical analysis indicator used to smooth out price fluctuations and identify trends in financial markets.
ADX	ADX stands for Average Directional Index, which is a technical analysis indicator used to measure the strength of a trend in financial markets. The ADX is a non-directional indicator, meaning it does not indicate whether the trend is bullish or bearish, but rather the strength of the trend.
MACD	The MACD is a technical analysis indicator that helps identify potential buy or sell signals in financial markets. It stands for Moving Average Convergence Divergence and works as a trend-following momentum indicator by showing the relationship between two moving averages of prices.
CCI	CCI stands for Commodity Channel Index, which is a technical analysis indicator used to identify potential overbought or oversold conditions in financial markets. The CCI is a momentum oscillator that shows the relationship between the current price, a moving average of prices, and standard deviations from the moving average.
MO	The Momentum Oscillator is a technical analysis indicator used to measure the rate of change in a financial instrument's price. It is a type of oscillator that oscillates around a zero line and shows whether the price of an asset is gaining or losing momentum.
RSI	The Relative Strength Index is a technical analysis tool that measures the strength of a financial instrument's price action. It oscillates between 0 and 100, functioning as a momentum oscillator to identify potential overbought or oversold conditions in the market.

This paper utilizes various financial indicators, including the current ratio, quick ratio, equity ratio, debt-to-equity ratio, inventory turnover, current asset turnover, fixed asset turnover, diluted earnings per share growth rate, return on net assets, gross operating margin, net operating margin, price-to-earnings ratio (P/E), price-to-net ratio (P/N), price-to-sales ratio (P/S), and consolidated leverage. These ratios are chosen because they are readily available in the company's financial statements and can directly impact its stock price. Table 2 provides further evidence of this relationship.

Table 2: Introduction of financial indicators.

Indicators	Description
Current Ratio	The current ratio is a financial measure that assesses a company's ability to pay off its short-term debts, due within one year, using its current assets. Current assets are those that can be converted into cash within the same timeframe.

DOI: 10.54254/2754-1169/45/20230275

Table 2: (continued).

Quick Ratio	The quick ratio measures a company's ability to pay off its short-term debts using its most liquid assets. It differs from the current ratio in that it excludes inventory and other assets that may not be easily converted into cash.
Equity Ratio	The equity ratio is a financial metric that gauges the percentage of a company's assets funded by its shareholders' equity. This ratio assesses a firm's financial leverage and reveals the extent to which shareholders, rather than creditors, own its assets.
Debt/Asset Ratio	The debt-to-asset ratio, or debt ratio, is a financial metric that indicates the percentage of a company's assets funded by debt. The formula is defined as $\frac{total\ amount\ of\ dept}{total\ amount\ of\ assets}$.
Inventory Turnover Ratio	The inventory turnover ratio is a financial measure that indicates how efficiently a company sells and replenishes its inventory within a given timeframe, typically one year.
Current Asset Turnover Ratio	The current asset turnover ratio gauges how effectively a company can generate sales revenue from its present assets. The formula is defined as $\frac{net\ salesof\ acompany}{average\ current\ assets}$.
Fixed Asset Turnover Ratio	The fixed asset turnover ratio is a financial metric used to assess how efficiently a company generates sales revenue from its fixed assets. The formula is defined as $\frac{net\ sales\ of\ acompany}{average\ fixed\ assets}$.
Diluted Earnings Per Share (EPS) Growth Rate	The diluted earnings per share (EPS) growth rate is a financial metric that indicates the percentage change in a company's diluted EPS over time. The formula is defined as $\frac{current\ period's\ diluted\ EPS - previous\ period's\ diluted\ EPS}{previous\ period's\ diluted\ EPS} \times 100$.
ROE	ROE is an acronym for Return on Equity, a financial ratio that gauges a company's profitability by dividing its net income by the shareholders' equity and expressing it as a percentage. In simpler terms, ROE indicates how much profit a company earns per dollar of shareholder equity.
Gross Profit Margin	The gross profit margin is a financial ratio used to determine a company's profitability. The formula is defined as $\frac{total\ revenue - COGS}{total\ revenue} \times 100$.
Net Profit Margin	The Net Profit Margin is a financial ratio used to assess the profitability of a business. It compares the net profit to revenue The formula is defined as $\frac{net\ profit}{total\ revenue} \times 100$.
P/E Ratio	The P/E Ratio evaluates the value of a company's stock. It is determined by dividing the current market price of the stock by its earnings per share (EPS) for the previous 12 months.
P/B Ratio	The P/B Ratio used to evaluate a company's valuation in relation to its book value. To calculate this ratio, divide the market price per share by the book value per share. The value can be determined by $\frac{total\ assets\ minus\ liabilities\ of\ a\ company}{outstanding\ shares}$

DOI: 10.54254/2754-1169/45/20230275

Table 2: (continued).

P/S Ratio	The P/S Ratio used to evaluate a company's valuation in relation to its revenue. The formula is defined as $\frac{the\ market\ capitalization\ of\ the\ company}{annual\ revenue}$ This metric is particularly useful for assessing companies with low or negative earnings.
Combined Leverage	Combined leverage is a financial metric that measures the effect of both operating leverage and financial leverage on a company's earnings before interest and taxes (EBIT). Operating leverage refers to the degree to which a company's fixed costs impact its earnings, while financial leverage refers to the degree to which a company's debt impacts its earnings.

The basic information of the technical indicators is shown in Table 3.

Table 3: Basic information of technical indicators.

Indicators	Mean	Std	Min	25%	50%	75%	Max
WMA	21.26	30.34	0.92	7.62	12.76	22.38	558.88
DEMA	21.29	30.55	0.94	7.63	12.72	22.47	598.67
ADX	25.24	10.55	3.89	17.49	23.08	30.82	100.00
MACD	0.00	1.51	-34.60	-0.26	0.01	0.25	30.69
MACDSIGNAL	0.04	1.36	-23.26	-0.21	0.03	0.26	27.36
MACDHIST	-0.03	0.68	-45.99	-0.09	-0.01	0.06	6.13
CCI	-7.86	110.32	-466.67	-97.41	-10.56	81.24	466.67
MO	0.00	4.03	-70.30	-0.79	-0.04	0.59	81.43
RSI	49.68	12.49	0.00	40.45	49.47	58.52	100.00

As can be seen, the overall indicators are relatively volatile, which may be due to the long period of data selection and the fact that this is a high-tech industry.

Before fitting the data, it is necessary to preprocess the data. Considering that the units of each indicator are different, each input variable is standardized to ensure that its magnitude is the same, which can avoid the performance loss caused by different units and allow the machine learning model to better identify the patterns in the data.

The basic information of the financial indicators is shown in Table 4 below.

Table 4: Basic information of financial indicators.

Indicators	Mean	Std	Min	25%	50%	75%	Max
Current Ratio	3.95	10.42	-23.73	1.48	2.22	4.12	741.90
Quick Ratio	3.32	9.81	-19.41	1.07	1.72	3.34	699.30
Equity Ratio	0.71	0.84	0.01	0.24	0.50	0.94	42.15
Debt/Asset Ratio	0.35	0.18	0.01	0.19	0.33	0.48	0.98
Inventory Turnover Ratio	150.12	13626.73	0.00	1.11	2.16	3.96	1643917.94
Current Asset Turnover Ratio	0.58	0.45	0.00	0.25	0.47	0.79	5.93
Fixed Asset Turnover Ratio	11.32	107.85	0.00	1.11	2.39	5.37	6558.08

DOI: 10.54254/2754-1169/45/20230275

Table 4: (continued).

Diluted Earnings Per Share (EPS) Growth Rate	0.76	7.10	-66.29	-0.47	0.00	0.56	263.00
ROE	0.06	0.05	0.00	0.02	0.04	0.08	0.71
Gross Profit Margin	0.29	0.14	-0.14	0.18	0.26	0.37	0.93
Net Profit Margin	0.12	0.13	0.00	0.05	0.09	0.15	5.82
P/E Ratio	105.13	791.67	0.00	33.29	51.49	88.69	52681.76
P/B Ratio	4.80	3.98	0.00	2.65	3.86	5.73	134.42
P/S Ratio	7.46	9.10	0.00	2.63	4.78	9.15	337.03
Combined Leverage	1.59	6.32	-159.16	1.00	1.17	1.55	621.72

As can be seen, the standard deviation of the Current Ratio, Quick Ratio, Inventory Turnover, Fixed Asset Turnover, Diluted EPS Growth Rate, P/E Ratio, and Consolidated Leverage is relatively large, which may be due to the presence of relatively large outliers in the data, thus increasing the standard deviation of the data. In such a case, one can focus on the median as well as the interquartile range. In addition to the standard deviation, the profitability of the industry as a whole can be seen from the return on net assets, the gross operating margin, and the net operating margin. From the data in the table, the result can be seen that the average return on net assets for the industry is 6%, which is still relatively low. The average gross operating margin is 29%, which is relatively high among all industries and in line with the characteristics of the industry.

4. Result

The models selected above were used to fit the data, and predictions were made on the test set, using f1, check-all rate, check-accuracy rate, and AUC as evaluation metrics, and the measured results are as follows (See Tables 5-7 and Figs. 1-3).

Table 5: The f1 score of each model.

Model	Technical Indicators only	Financial Indicators only	all
LogisticRegression	0.76	0.76	0.76
SVC	0.76	0.76	0.77
RandomForest	0.89	0.86	0.89
XGBoost	0.87	0.84	0.89

Table 6: The recall rate of each model.

Model	Technical Indicators only	Financial Indicators only	all
LogisticRegression	0.95	0.99	0.95
SVC	0.93	0.98	0.91
RandomForest	0.93	0.95	0.94
XGBoost	0.91	0.89	0.91

DOI: 10.54254/2754-1169/45/20230275

Table 7: The precision rate of each model.

Model	Technical Indicators only	Financial Indicators only	all
LogisticRegression	0.63	0.62	0.63
SVC	0.65	0.62	0.66
RandomForest	0.85	0.79	0.85
XGBoost	0.84	0.80	0.86

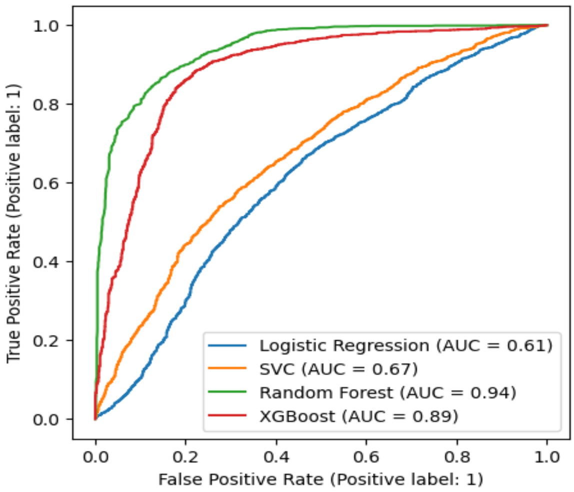

Figure 1: The ROC curve of only technical indicators in the models.

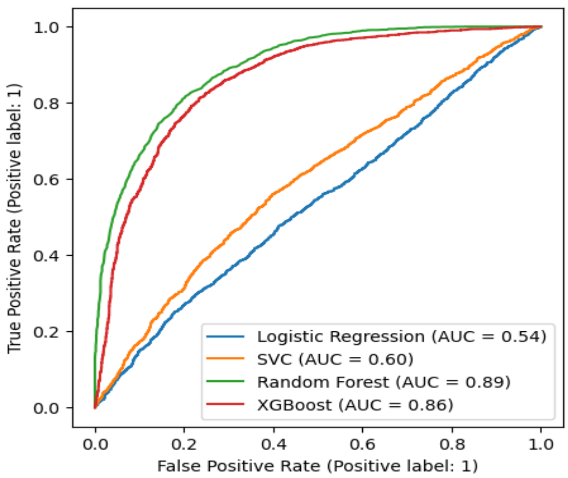

Figure 2: The ROC curve of only financial indicators in the models.

DOI: 10.54254/2754-1169/45/20230275

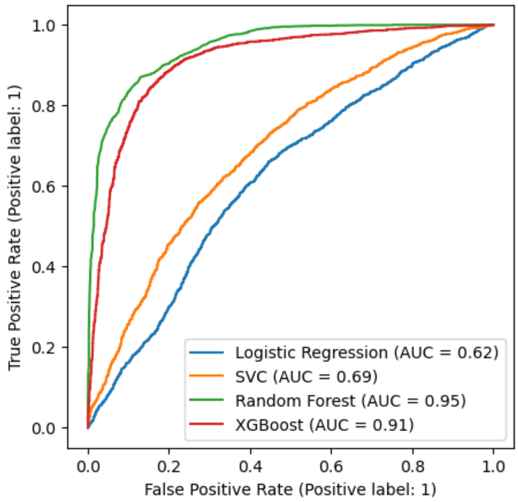

Figure 3: The ROC curve of all indicators model.

It can be seen from Table 5 and the figures above that the results of Random Forest and XGboost are significantly improved compared to Logistic Regression and Support Vector Machine. Further observation from Tables 6 and 7 shows that the improvement of the f1 indicator is mainly due to the improvement of the accuracy of the control. In addition, both technical and financial indicators can achieve better results in predicting performance one day after the announcement of financial statements. This indicates that in addition to stock price information, financial indicators also deserve investors' attention, and financial indicators have informational value. In addition, the overall performance of the model is slightly improved after considering both technical and financial indicators. This indicates that the model achieves better performance as financial information provides information beyond stock price information.

5. Conclusion

The research in this paper is about using financial information and technical indicator information for stock price forecasting. This paper selects specific industries in the Guotaian database, selects the financial indicator information of each listed company in the industry, obtains the stock price data of these companies from Yahoo Finance, calculates the corresponding technical indicators, and calculates whether the stock price increases on the second trading day after the company's earnings report is released. The machine models used in this paper are Logistic Regression, Support Vector Machine, Random Forest, and XGBoost, and the performance of the models using only technical indicators, only financial indicators, and both are compared to determine the degree of contribution of technical indicators, financial indicators, and both together to stock price prediction. It is found that the model using only technical indicators outperforms the model using only financial indicators, but the difference is not significant. The models that use both perform slightly better than the models that use indicators alone. This suggests that financial indicators can be used to predict stock prices. And financial indicators do play a role in the stock price forecasting model. Of course, there are some shortcomings in this study, for example, more models could be tried to see the importance of financial indicators in stock price forecasting. In addition, no further model tuning was done in this study, and all models used the initial settings, which means that there is room for further model improvement.

DOI: 10.54254/2754-1169/45/20230275

References

[1] *Liu, X.Y., Cheng, X. M.: Stock price prediction model based on the financial index. Journal of Beijing University of Information Science and Technology (Natural Science Edition) (1), 96-100 (2022).*
[2] *Wu, Y.X., Wen, X.: Short-term stock price forecasting based on ARIMA model. Statistics and Decision Making 23, 83-86 (2016).*
[3] *Lin, Z.: Modelling and forecasting the stock market volatility of SSE composite index using GRACH models. Future Generation Computer Systems 79, 960-972 (2018).*
[4] *Zhang, G.S., Zhang, X.D.: A differential information-based ARMAD-GARCH stock price forecasting model. Systems Engineering Theory and Practice (05), 1136-1145 (2016).*
[5] *Lin, X., Zhu, S.D.: Attention-mechanism-based LSTM model for stock price predicting. Journal of Chongqing University of Technology and Business (Natural Science Edition) (02), 75-82 (2022).*
[6] *Ohlson, J. A.: Earnings, book values, and dividends in equity valuation. Contemporary Accounting Research 11(2), 661-687 (1995).*
[7] *Bernard, V. L., Stober, T. L.: The nature and in amount cash flows of information and accruals. The Accounting Review 64(2), 624-652 (1989).*
[8] *Beaver, W. H., Roland, E. D.: Earnings expectations and the behavior of security prices. The Accounting Review 47(2), 320-332 (1972).*
[9] *Sloan, R. G.: Do stock prices fully reflect information in accruals and cash flows about future earnings? The Accounting Review 71(3), 289-315 (1996).*
[10] *Habib, A.: Value relevance of alternative accounting performance measures: Australian evidence. Accounting Research Journal 23(2), 190-212 (2010).*