

# Predictive Analysis of AAPL Stock Trend by Random Forest and K-NN Classifier

Wenxi Qin\*

Department of Social Science, University of California Irvine, California, US

\*Corresponding author: Wenxiq@uci.edu

**Abstract.** While perfection in stock market prediction is impossible to accomplish, minimizing the investment risk utilizing data mining classifier to forecast the Stock Market trend becomes one of the most popular fields to research. This paper analysis and forecast AAPL 's daily stock trends by utilize two most frequently use data mining algorithm K-Nearest Neighbor and Random Forest Classifier. By comparing the accuracy of the two model to finds out which model would best estimate the AAPL's stock price. Both Algorithm utilize simple rolling average as the predictors to increase the accuracy of the prediction. The result points out that random forest classifier has 61% of the total accuracy which is greater than K-Nearest Neighbor (KNN) with only 53% of accuracy in total for forecasting the AAPL stock trend. Random Forest Classifier have a better performance in predicting the AAPL's stock market compared to KNN, however it may show a different result when it comes to other stock data.

**Keywords:** APPLE; Random Forest; KNN.

## 1. Introduction

Stock prices are not random, in fact, risk of investment can often reduce by utilize the market predicting algorithm. By the utilization of the data market predicting algorithm, investor discovers the pattern between data and make prediction [1]. Traders consistently seek stock that exhibit an upward trend, emphasizing the importance of accurately predicting the stock's direction [2]. Studies have shown Stock trend does not limit on pure technical data, but also fundamental data which includes other features such as Inflation, short-term interest rate, firm size, market equity etc [3]. Finding an efficient way to predict the stock prices are often complicate task to accomplish since they are non-linear, time carrying, dynamic and affect by multiple factors [2]. In addition, it is shown that it would be better to predict the stock by treat it as a classification problem instead of a regression problem [4].

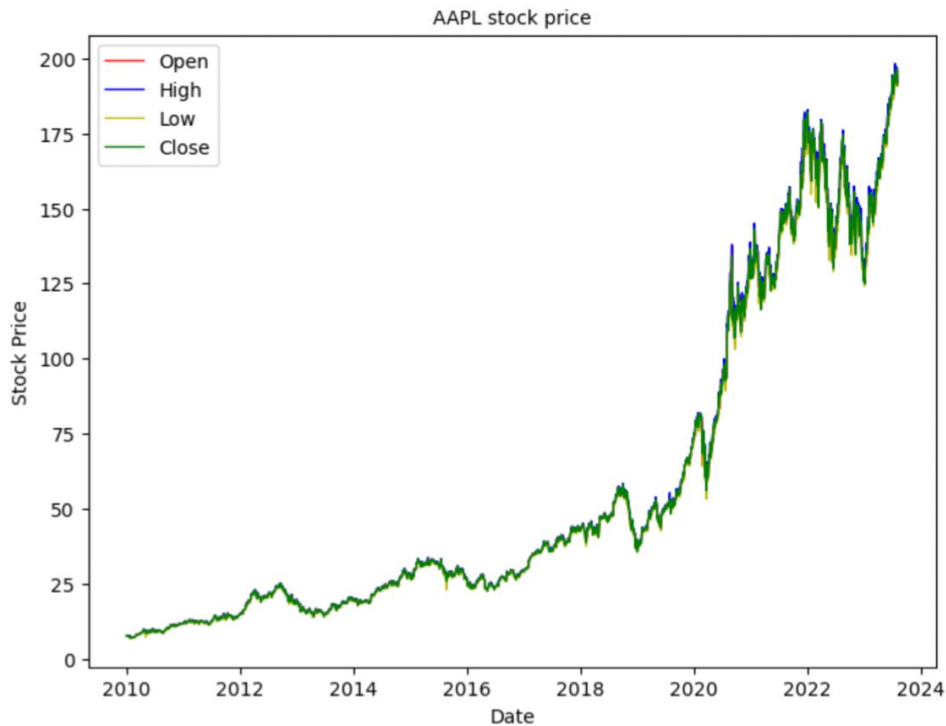
Classification can be used in multiple fields such as bioinformatics, loan risk prediction, medical diagnosis, and Market direction prediction [3]. As two the foremost classifier producing high accuracy, Random Forest, K-Nearest Neighbor often employs in Forecasting Market trend [5]. While K-NN as one of the most stable algorithms with high accuracy [6], it is shown to be a a time consume algorithm, which it runs significantly slow as the increase of numbers of data [7]. Random Forest have similar disadvantages, it is also time-intensive in generating forest during training [8]. However, some study suggests that the performance of the random forest classier turns out to be better than other forms of classifier since it is robust against overfitting and bias [9-10].

In fact, there have been many similar studies question the accuracy of the prediction of KNN and random forest classifier, which motivates the author to make a comprehensive investigation on this issue. In this paper, the famous KNN and random forest classifier are adopted to make prediction of the stock price trend of AAPL.

## 2. Data

This paper tackles the issue of stock price forecasting with different data mining algorithm. The predictors are based on the ratio between the closing price of today and the rolling average of the closing price over different time (2, 5, 60, 250, 1000 trading days), as well as the average trend over

these same periods. Additionally, there is a Target column where a value of 1 represent an upward trend in the market and 0 represent the opposite. The AAPL stock data are download form Yahoo Finance which include from 2010-01-06 to 2023-08-04 (See Fig. 1). The data contains about 3418 trading days about AAPL Market price. The training set of the data includes all the data except for the last 100 trading days which utilize as testing set to test the accuracy of the data.



**Fig.1** Visualization of AAPL Stock Price

### 3. Methodology

This Paper would use two classifiers to examine the stock trend of the AAPL’s stock Market, which is K-NN and random forest classifier to compare the accuracy of the classifier in the given stock market. The accuracy rate will be evaluated by the confusion matrix.

#### 3.1. K-Nearest Neighbor Classifier

K-Nearest Neighbor Classifier is a simplest yet highly accurate data mining algorithm. This is a algorithm that does not need upfront construction of function. K-NN finds the nearest neighbor by utilize the Euclidean distance metrics. In this algorithm, N represent the number of features. The equation of the metrics is:

$$d(x, y) = \sum^N x^2 y^2 \tag{1}$$

In this paper, K-NN would be using to solve a binary classification problem to predict the Column Target, whether the closing price tomorrow would show an uptrend in compared to today’s closing price.

The model predicts the trend of the AAPL Market by looking at K=3 nearest neighbor and utilize the predictors that based on the ratio between closing price of today and the rolling average of the closing price over different time period (2, 5, 60, 250, 1000 trading days), as well as the average trend over these same period. with the Target, where 0 represent downtrend and 1 represent uptrend to train the data.

### 3.2. Random Forest Classifier

Random Forest predicts the stock market trend by create a forest with a substantial numbers of decision trees for training. During classification problem, the algorithm takes the majority vote. Instead, during regression problem, the algorithm takes the average. The algorithm has a better performance in case of classification problem compared to regression problem. In addition, Random Forest classifier is less bias because each of the trees build on a random select sample, each tree is grown randomly.

### 3.3. Predictors

The model uses 100 estimators and min sample split to prevent overfitting the model and use the predictors that based on the ratio between closing price of today and the rolling average of the closing price over different time period (2, 5, 60, 250, 1000 trading days), as well as the average trend over these same period. In addition, the Target, where 0 represent downtrend and 1 represent uptrend to train the data. The process means to find the relationship between the predictors and the target to forecast the trend of the AAPL stock Market.

### 3.4. Method of Evaluation

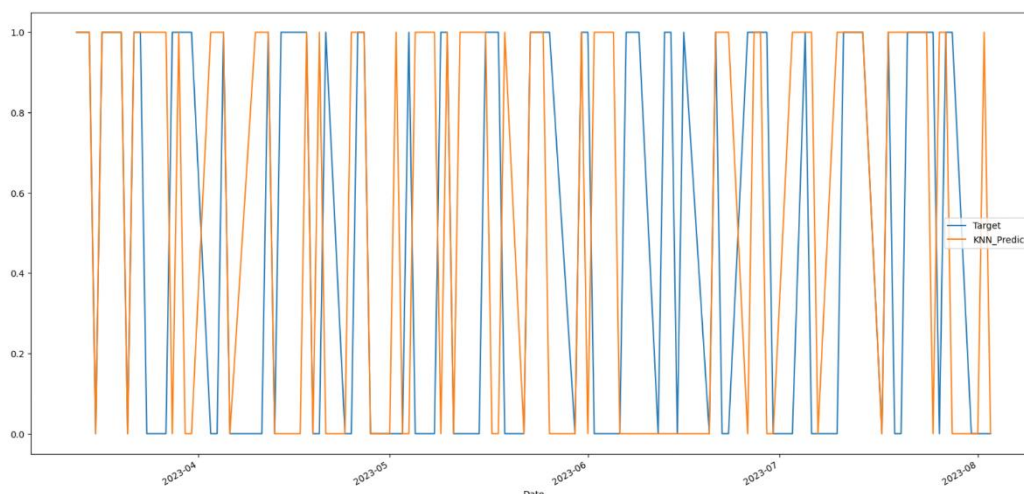
The performance of the algorithm will be assessed utilize the Confusion Matrix. The matrix includes True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The total accuracy would be based on the confusion matrix and calculate as the function below.

$$\text{Total Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \quad (2)$$

## 4. Result

### 4.1. KNN Result

The result of the prediction by KNN Classifier is shown in Fig. 2:



**Fig. 2** Visualization of AAPL Stock trend prediction using KNN

Table 1 shows that the true positive and negative value is 31 and 22, the false negative and false positive value is 22 and 25, which gives a total accuracy of 0.53. Which shows a low accuracy in predicting the market trend of AAPL.

**Table 1.** KNN Accuracy and Error Matrixes table

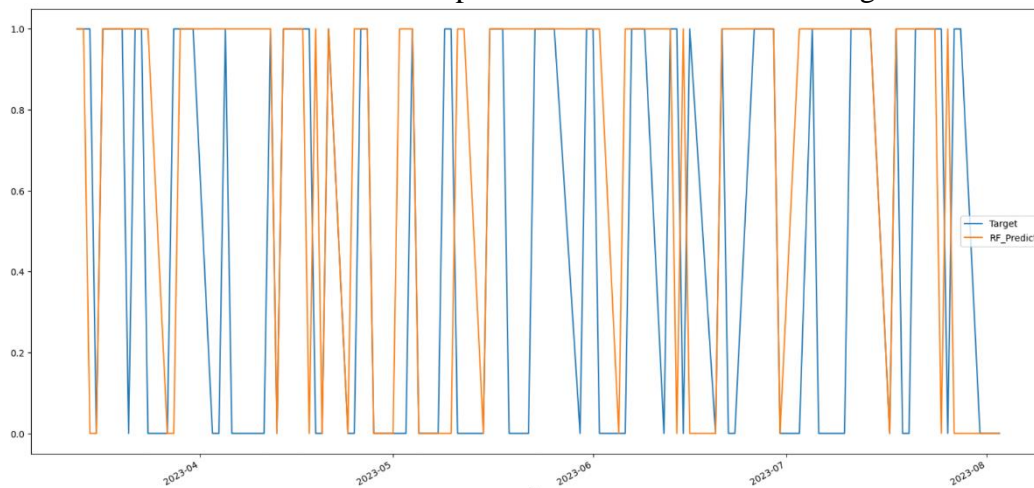
	Predict=0	Predict=1	Total
Target=0	22	25	47
Target =1	22	31	26
Total	44	56	

The KNN predict 56 times of the uptrend of the AAPL Market and only 31 times the market went up, 25 times when the KNN states the Market when up while it when down. So, the sensitivity of the result is about 0.58.

The KNN predict 44 times of the down trend of the data, only 22 times when it said down trends the market when down, and 22 times it states is a down trend the market went up, so the specificity of the result is about 0.47.

**4.2. Random Forest Classifier Prediction**

The result of the random forest classifier prediction is shown below in Fig. 3:



**Fig. 3** Visualization of AAPL Stock trend prediction using Random Forest

Table 2 shows that the true positive and negative value is 43 and 18, the false negative and false positive value is 10 and 43, which gives a total accuracy of 0.61.

**Table 2.** Random Forest Accuracy and Error Matrixes table

	Predict=0	Predict=1	Total
Target=0	18	29	47
Target =1	10	43	14
Total	28	72	

The Random Forest predict 72 times of the uptrend of the AAPL Market and only 43 times the market went up, 29 times when the Random Forest said the Market when up while it when down. So, the sensitivity of the result is about 0.81.

The Random Forest predict 28 times of the down trend of the data, only 18 times when it said down trends the market when down, and 10 times it saids is a down trend the market went up, so the specificity of the result is about 0.38.

**4.3. Compare and discuss the result.**

The result shows random forest classier has overall higher performance in prediction of the market trend, since the total accuracy of the random forest classification prediction is about 0.61 while the total accuracy of the K-NN classification prediction is about 0.53. In addition, random forest classification has higher performance on predicting the uptrend, since the sensitivity of the random

forest classification is 0.81 while K-NN has a lower sensitivity of 0.58. However, the result shows that the K-NN has higher performance on predicting the stock down trend, the specificity of the K-NN prediction is about 0.47 which is higher than the specificity of random forest classification 0.38. Both algorithms utilize only the rolling average of the closing price as predictor of the algorithm. The predictor is more focus on the pure technical data instead of the fundamental data of the stock.

## 5. Conclusion

The paper compares the result of the prediction of the two-classification method, KNN and Random Forest Classification. Random forest classification has higher total accuracy and better performance on uptrend prediction, while KNN have higher performance on the down trend prediction on AAPL stock trend prediction. However, the total accuracy of prediction of both algorithm is still low. The change of stock price does not only limit to pure technical data, but also fundamental data such as firm size, market equity, inflation, and short-term interest rate etc. The limitation of the study maybe totally ignores these factors. As a result of that, both algorithms may be improved by adding fundamental data such as compose activity and the situation of the market. By adding more fundamental data to the algorithm, the training of the data gaining more information about to identify the relationship, which may result in improvement of the precision of both algorithms. Future studies may focus on improving the accuracy of the algorithm by identify the most relate factors of training the algorithm, and other feature engineering method.

## References

- [1] Imandoust, S. B., & Bolandraftar, M. Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International journal of engineering research and applications*, 2013, 3(5): 605-610.
- [2] Tan, T. Z., Quek, C., & Ng, G. S. Biological brain-inspired genetic complementary learning for stock market and bank failure prediction 1. *Computational intelligence*, 2007, 23(2): 236-261.
- [3] Subha, M. V., & Nambi, S. T. Classification of Stock Index movement using k-Nearest Neighbours (k-NN) algorithm. *WSEAS Transactions on Information Science & Applications*, 2012, 9(9): 261-270.
- [4] Khaidem, L., Saha, S., & Dey, S. R. Predicting the direction of stock market prices using random forest. 2016. arXiv preprint arXiv:1605.00003.
- [5] Boateng, E. Y., Otoo, J., & Abaye, D. A. Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. *Journal of Data Analysis and Information Processing*, 2020, 8(4): 341-357.
- [6] Alkhatib, K., Najadat, H., Hmeidi, I., & Shatnawi, M. K. A. Stock price prediction using k-nearest neighbor (kNN) algorithm. *International Journal of Business, Humanities and Technology*, 2013, 3(3): 32-44.
- [7] Halvani, O., Steinebach, M., & Zimmermann, R. Authorship verification via k-nearest neighbor estimation. *Notebook PAN at CLEF*, 2013.
- [8] Abraham, R., Samad, M. E., Bakhach, A. M., El-Charani, H., Sardouk, A., Nemar, S. E., & Jaber, D. Forecasting a stock trend using genetic algorithm and random forest. *Journal of Risk and Financial Management*, 2022, 15(5): 188.
- [9] Ansary, M. S., & Brinto, A. Prediction of Stock Price Direction with Trading Indicators using Machine Learning Techniques. In *2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering*. 2022: 1-4.
- [10] Breiman, L. Random forests. *Machine learning*, 2001, 45: 5-32.