Stock Trading Using k-NN and Technical Analysis: A Case for Turkish Stocks

Ömer Mustafa Atagül, Ahmet Erdem Çetintaş, Zita Engels, Nida Nalbant, Ece Rende

Abstract

In the following, we are computing a model to predict the future daily returns of fifteen Turkish stocks by employing the technical analysis approach of Teixeira and Oliveira (2010).
For each stock, we first analyzed the technical features concerning daily price and volume data in order to use them as inputs for the k-NN algorithm. To compare our results, we have used the buy-and-hold strategy as a benchmark. Our model's cumulative return is less than the benchmark's, though we receive quite high results in terms of classifier accuracy and precision. Also, we observe, that there is no specific k-value that brings the most cumulative returns compared to the benchmark model. Therefore, we define our model as not feasible and propose further investigation in machine learning models for stock market prediction.

1 Introduction

Stock price prediction is an interesting and challenging task due to its dynamic and nonlinear information sources, which has been hard to capture for traditional economic models (Teixeira and Oliveira, 2010). With machine learning methods there are new possibilities to compute more accurate prediction models than before. We wanted to model a trade prediction system for fifteen stocks from Borsa Istanbul (BIST30 companies) and therefore employed the nearest neighbor (NN) classifier.

In order to find a feasible approach, we viewed at previous literature, capturing machine learning methods on exchange market analysis. Alkhatib et. al. (2013) employ a k-NN algorithm and a nonlinear regression approach to predict closing prices for a Jordan stock sample, which consists of 200 records for each of the six major companies in the country. By evaluating their results with the smallest root mean square error, they find k-NN with k=5 to be robust and highly accurate.

Theofilatos et. al. (2012) model a trading system for the EUR/USD exchange rate, employing five different machine learning techniques for predicting one day ahead movement with only autoregressive terms as input. Also, they vary their results, once modeling without and once with transaction costs, which is quite important, considering that their models use a high rate of trade. They find, that Support Vector Machines (SVM) and Random Forest bring most accurate results in terms of annualized returns and Sharpe ratio.

Lastly, we consider the approach of Teixeira and De Oliveira, whose analysis we will later try to resemble: They argue that, following the hypothesis of Efficient Markets, the best choice in the market would be to follow the buy-and-hold strategy, which implies to buy and never sell securities. To test this hypothesis, they compute a stock trading model and evaluate its feasibility using buy-and-hold profitability as a benchmark. Next to the employment of k-NN classifier they use different technical analysis tools, which are used for stock valuation, such as short- and long-term Simple Moving Averages (SMA) for either 10 or 21 periods, Relative Strength Index (RSI) filter for 14 periods or stop-loss and stop-gain. To train their model, they use the history of daily stock closing prices and volumes of 15 real stocks from Sao Paulo Stock Exchange over April 1998 till March 2009, including transaction costs. Even though the precision of the classifier is relatively low (<50%), the method performs better in terms of profitability, than the benchmark strategy for 12 out of 15 stocks, and therefore can be seen as a feasible trading system under real market conditions.

For our Analysis we will implement a similar approach as Teixeira and Oliveira: We will combine k-NN classifier with the same technical tools, using also similar values. Even, if Theofilatos, Likothanassis, and Karathanasopoulos choose different machine learning techniques, the nearest neighbor is a simple and also fast method which, Teixeira and Oliveira argue, is an important feature for stock exchange prediction.

The following parts will be organized as followed: Section 2 will explain the technical analysis tools we introduced, similar to Teixeira and De Oliveira's approach. Section 3 will illustrate how the k-NN method is implemented for the Stock Return Sign Prediction, covering Data, Statistics, and the computed Model. Section 4 explains the found results and section 5 concludes.

2 Technical Analysis

Technical analysis examines the market action by using charts for the prediction of future price movements (Murphy, 1999). Market action consists of three principal sources that are available to the technician: price, volume, and open interest. According to the Efficient Market Hypothesis (EMH) all available information is reflected by the prices. This rule holds in technical analysis with the belief of the technician that all past information is reflected in the prices and all the new information will be reflected in those prices immediately. Therefore, analyzation of the history of the prices is the necessary action to be made by the technician. This project focuses on the information on the prices and volume of trades. With these information technical indicators are constructed. Technical indicators are mathematical calculations that using the information of price or volume that came from the data of a security for identifying the movement of those prices (Teixeira and Oliveira, 2010). Moving averages and the oscillators are the most commonly used technical indicators. The moving average is a trend following device with smoothing the price movements by averaging the price data. In comparison, oscillators are used to identify the momentum.

As the paper suggests, we are applying mainly four indicators in this project. Moving averages, Relative Strength Index (RSI), Stochastics, and Bollinger bands. There are so many technical indicators in the market but the paper we examine chooses these indicators considering two main arguments. The first one is if the technical indicator is practically widely used by technicians. And the second one is if the technical indicators are coming from different categories like moving averages and the oscillators.

2.1 Moving Averages

The paper uses the Simple Moving Average (SMA), which is calculated by the arithmetic mean of a specified period is the most used type by technicians. For each day, the calculation is made and the smooth curve is built with the calculations. It becomes easier with this curve to identify the trend of the prices. The SMA formula for n days period is (1).

$$SMA(t) = \frac{1}{n}\sum_{i=t-n}^{t} x(i) \tag{1}$$

2.2 Relative Strength Index (RSI)

The RSI is the most popular momentum indicator that measures the magnitude of recent price changes to identify overbought or oversold conditions in a stock. The RSI is an oscillator which is a line graph that moves between two extremes and the RSI index ranges between 0 to 100. There are two critical points in the interpretation of the RSI. When the RSI index is 70 or above, it means that a security is becoming overvalued which is a bearish signal. The RSI generates the opposite signal that is a bullish signal when the RSI is 30 or below since it means that a security is undervalued. RSI is calculated by the equations from (2) to (4).

$$RSI = 100 - \frac{100}{1+RS}, \quad \text{where } RS = \frac{avgGain}{\text{avg}Lost} \tag{2}$$

$$avgGain = (\text{total of gains during past } n \text{ periods})/n \tag{3}$$

$$avgLost = (\text{total of losses during past } n \text{ periods})/n \tag{4}$$

2.3 Stochastics

A stochastic oscillator is another momentum indicator with the aim to identify where the most recent closing price is in relation to the price range for a time period. This indicator consists of two lines, the fast (%K) and the slow (%D), that are calculated via the (5) and (6) respectively.

When the closing levels that are persistently close to the top of the range imply accumulation which shows the buying pressure. When the closing price is close to the bottom of the range imply distribution which shows the selling pressure.

$$\%K = 100 \times \frac{recentClose - lowestLow(n)}{highestHigh(n) - lowestLow(n)} \quad (5)$$

$$\%D = 3 - period\ moving\ average\ of\ \%K \quad (6)$$

2.4 The Bollinger Bands

As another mostly used technical analysis tool, the Bollinger bands consist of three bands to compare the relative price levels and the volatility over a period time; an SMA in the middle, an upper band which is formulated as an SMA plus a number of standard deviations, and a lower band formulated as an SMA minus a number of standard deviations. Bollinger bands can be used as a signal generator for significant moves when it is used with price action.

3 Using k-NN to Stock Return Sign Prediction

3.1 k-NN

K-nearest neighbor classification is known as its simple implementation as a machine learning algorithm (Aha et al., 1991). Its aim is to classify a new object by measuring similarities between other available objects, as known as its "neighbors". A distance function, such as Euclidean, is used in order to measure the similarity. The nearest neighbors of the main object is determined in order to taken into account in the classification process. In this case, k-value represents the number of neighbors that the algorithm uses to assign the object's class. According to the algorithm, the object belongs to the majority of its k nearest neighbors. When k=3, the object is evaluated by the majority vote of the 3 nearest neighbors. If k=1, since there will be only one neighbor, the object is classified as its nearest neighbor. Simply its implementation works as follows:

- Determining the value of k
- Computing the distance between the k nearest neighbors of the object
- Assigning the class label as the majority of the k nearest neighbors (Alkhatib et al., 2013)

3.2 Features

3.2.1 Data

Fifteen stocks from Borsa Istanbul (from BIST30 companies) were chosen to be used in our analysis. (companies introduced in Figure 1) The data we used consist of daily open, high, low, close prices and volume of each stock. We have taken the time interval from 01/01/2009 to 31/12/2018.

Ticker	Company Name	Ticker	Company Name	Ticker	Company Name
AKBNK	Akbank	GARAN	Garanti Bankası	TAVHL	TAV Havalimanları
ARCLK	Arçelik	KCHOL	Koç Holding	TCELL	Turkcell
DOHOL	Doğan Holding	KOZAA	Koza Madencilik	THYAO	Türk Hava Yolları
EREGL	Ereğli Demir Çelik	SAHOL	Sabancı Holding	TKFEN	Tekfen Holding
FROTO	Ford Otosan	SISE	Şişe Cam	TTKOM	Türk Telekom

Figure 1. Company names

For each stock and for each day, we have computed technical analysis indicators from daily prices and volume data which were mentioned above. Then, these technical analysis indicators were used as inputs to k-NN algorithm. The complete list of technical analysis indicators we used was shown in Figure 2.

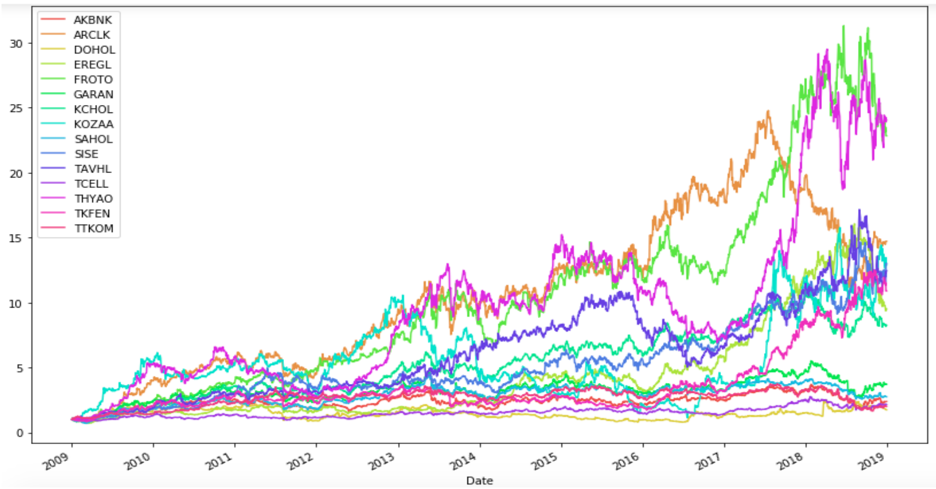
$$
\begin{aligned}
&I_1 = RSI(t)\\
&I_2 = (x(t) - bollinger_{upper})/bollinger_{upper}\\
&I_3 = (x(t) - bollinger_{lower})/bollinger_{lower}\\
&I_4 = \%K(t)\\
&I_5 = \%D(t)\\
&I_6 = \%K(t) - \%K(t-1)\\
&I_7 = \%D(t) - \%D(t-1)\\
&I_8 = (x(t) - x(t-1))/x(t-1)\\
&I_9 = (PMA_s(t) - PMA_s(t-1))/PMA_s(t-1)\\
&I_{10} = (PMA_l(t) - PMA_l(t-1))/PMA_l(t-1)\\
&I_{11} = (PMA_s(t) - PMA_l(t-1))/PMA_l(t-1)\\
&I_{12} = (x(t) - PMA_l(t))/PMA_l(t)\\
&I_{13} = (v(t) - v(t-1))/v(t-1)
\end{aligned}
$$

Figure 2. Technical analysis indicator list.

We have used the parameter values of technical analysis indicators the same as used in Teixeira and De Oliveira (2010). Thus, the short-term moving average was 10 days and the long-term moving average was 21 days. In Bollinger bands, we have used 8 days simple moving average and 2 standard deviations. To calculate stochastics %K and %D, we have used 14 for fast K and 3 for fast D. For RSI, we have used 14.

3.2.2 Statistics

There were some missing values in our data. We have overcome them by forward filling the missing data, which mean that we have filled the missing values with the previous day's value. In this subsection, we present some important graphs and statistics in our data.



	Cumulative Return
THYAO	24.044776
FROTO	22.834862
ARCLK	14.716981
SISE	13.000000
KOZAA	12.892857
TAVHL	12.458333
TKFEN	10.887701
EREGL	9.500000
KCHOL	8.255814
GARAN	3.702326
SAHOL	2.744444
AKBNK	2.384615
TCELL	2.150177
TTKOM	2.000000
DOHOL	1.741379

Figure 3. Cumulative returns between 2009 -2018

Initial close prices are set to 1 to see the cumulative return. As the statistics show that, if we bought THYAO on 01/01/2009 with 1000 TL, we would have 24045 TL at the end of 2018.

	AKBNK Return	ARCLK Return	DOHOL Return	EREGL Return	FROTO Return	GARAN Return	KCHOL Return	KOZAA Return
count	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000
mean	0.000601	0.001306	0.000602	0.001107	0.001485	0.000785	0.001022	0.001607
std	0.022627	0.021865	0.027773	0.020554	0.021956	0.022990	0.019120	0.034675
min	-0.092920	-0.102226	-0.203252	-0.107383	-0.143986	-0.132251	-0.088843	-0.200000
25%	-0.013051	-0.010984	-0.013514	-0.010444	-0.010149	-0.013176	-0.009734	-0.016035
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	0.013423	0.012849	0.012658	0.013430	0.012842	0.014328	0.012549	0.015259
max	0.109848	0.194444	0.195402	0.109589	0.156028	0.132050	0.108374	0.200000

	SAHOL Return	SISE Return	TAVHL Return	TCELL Return	THYAO Return	TKFEN Return	TTKOM Return
count	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000	2516.000000
mean	0.000610	0.001241	0.001254	0.000462	0.001539	0.001190	0.000445
std	0.020438	0.021032	0.022400	0.017750	0.023439	0.021953	0.018396
min	-0.120909	-0.106952	-0.173520	-0.112481	-0.150594	-0.136461	-0.110476
25%	-0.010844	-0.011628	-0.010609	-0.009259	-0.012014	-0.011073	-0.009662
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	0.012076	0.013897	0.013409	0.010251	0.014657	0.013986	0.010791
max	0.108280	0.083333	0.099145	0.109402	0.122137	0.165316	0.101777

Figure 4. Statistics of daily returns

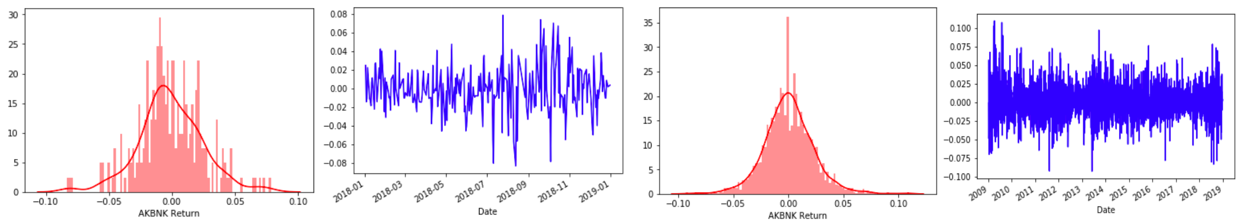


Figure 5. Daily returns of AKBNK for one year and for the whole period

Daily returns seem to fluctuate around 0 since the mean of all daily returns are close to 0. Also, a close look at the daily returns of AKBNK last year and through the whole period. It is seen that daily returns fluctuate rapidly around 0 through time. Likewise, in histograms, gathering around 0 is observed.

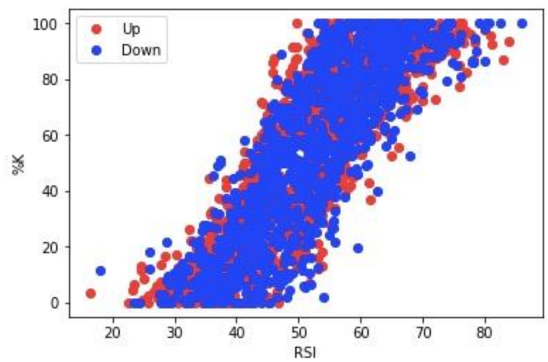


Figure 6. Data distribution

RSI on x-axis and %K on y-axis are plotted and the returns are classified as up and down in Figure 6. It is seen that the data is tangled and difficult to predict the price trend. However, we tried to predict it by employing k-NN algorithm to many technical analysis indicators.

3.3 Model

Our main objective in this analysis is to predict the sign of future daily returns. Each day we have calculated the technical analysis indicators which are our inputs to k-NN algorithm and our target variable is next day's return. We have classified returns as -1 if the close price decreased the next day and as 1 if the close price is constant or increased the next day.

We have carried out the walk-forward analysis to test the model as done in Teixeira and De Oliveira (2010). To do this, we have split the data set to 10 subsets corresponding to 10 yearly intervals. After that, we have trained the data for three successive years and tested the model with the next year. Then, we continued this process until the end of the subsets. In the end, we have tested seven subsets. The diagram of this process is pictured in Figure 7. We have used confusion matrix, accuracy and precision metrics and ROC curve to analyze our k-NN algorithm results.

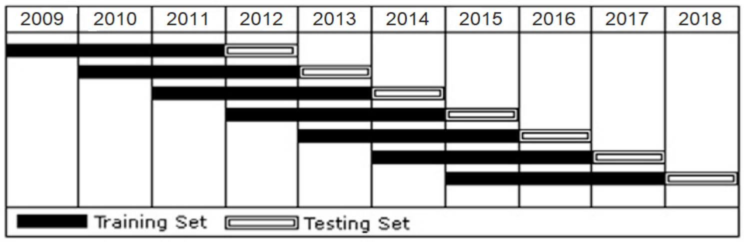


Figure 7. Walk-forward analysis

After predicting the next day's return, we have bought the stock if the predicted return was 1 and we have sold the stock if the predicted return was -1. The transactions costs were ignored. Then, we have calculated cumulative returns of the trades both at the end of each subset and at the end of whole data according to buy and sell orders for each stock. We have used the buy-and-hold strategy as a legitimate benchmark to compare the k-NN strategy. Therefore, we have also calculated the cumulative returns of buy-and-hold strategy for each stock both at the end of each subset and at the end of the whole data. Finally, we have computed five performance and risk metrics to compare the two strategies, which are CAGR (Compound Annual Growth Rate), volatility, maximum drawdown, Sharpe ratio and Calmar ratio.

4 Results

4.1 Evaluation Metrics

The evaluation metrics measure the classifier ability to identify classes correctly in the classification process. In machine learning applications, one of the way to measure classifier performances is accuracy. Accuracy is found by dividing the total number of correctly predicted instances by the total number of instances. In addition to that, we also used another measure to evaluate the classier performance which is precision, and before defining the precision, we will briefly examine a confusion matrix which shows correct and incorrect predicted instances in each class according to the actual class labels.

Actual/Predicted as	Positive	Negative
Positive	tp	Fn
Negative	fp	Tp

$$\text{precision} = \frac{tp}{tp + fp}$$

Figure 8. Confusion matrix for the two-class classification

In Figure 8, in the first table, there is a confusion matrix for the two-class classification task. In this matrix, tp and fp show true positive and false positive counts respectively. fn is false negative, and tn is true negative counts. The formula explains the precision. It shows how often the classifier predicts correctly when the prediction is positive. We examined and evaluated these metrics for each of all stocks to see the classifier ability to identify classes accurately.

STOCK	Accuracy	Precision	Difference to Buy&Hold
AKBNK	51.39%	53.23%	23.46%
ARCLK	51.43%	53.35%	-57.98%
DOHOL	60.33%	65.10%	-26.17%
EREGL	55.49%	58.43%	114.27%
FROTO	53.92%	56.58%	-17.05%
GARAN	53.78%	54.61%	402.37%
KCHOL	51.16%	53.55%	-57.41%
KOZAA	49.23%	49.90%	-98.34%
SAHOL	50.66%	52.30%	-66.60%
SISE	53.07%	56.13%	-29.22%
TAVHL	52.36%	54.58%	33.99%
TCELL	51.05%	55.06%	-49.39%
THYAO	52.24%	55.13%	-44.39%
TKFEN	49.03%	52.65%	-93.30%
TTKOM	51.62%	53.34%	2.60%

Figure 9. Evaluation metrics

As you can see from the above table, we also added to the table, our model's cumulative return difference to buy-and-hold benchmark strategy, to see whether there is a positive relationship between better-performing evaluation metrics and our model's return. Yet, even though the best performing evaluation metrics belong to stock DOHOL, Dogan Group, our model's cumulative return brings 26.17% less than the benchmark's. Nevertheless, on average, our classifier accuracy is around 52%, and the precision is around 55%, which are good numbers to have as a classifier. We also examined the k-parameter value, for each of all stocks up to 50, and its relation with an error rate and cumulative return difference with the benchmark strategy.

There is a general trend in increasing k-value and decreasing error rate for most of the stocks we have examined. However, we could not find any common the 'best' k-value to implement and to have proper results in our model. Another relationship that we examined is between k-value and our model's cumulative return difference to the benchmark model. For example, in terms of cumulative return, the best k-value for KCHOL is 5, and for THYAO it is 40. (For all stocks, see Appendix) There is no, again, any specific k-value that brings the most positive cumulative return compared to the benchmark model.

Receiver operating characteristic (ROC) and area under the ROC (AUC), is a performance measurement for classification. ROC is a probability curve and the area under the curve represents degree or measure of separability. It basically tells us how capable our model is to distinguish between classes. Therefore, the higher the area, the better the model. When AUC is 0.5 it means the model has no separation capacity whatsoever. We did the ROC analysis for all of 15 stocks and the graph below belongs to GARAN, Garanti Bank, which is the best AUC we get from the model. Thus, we cannot say that our model k-NN is good at distinguishing positive and negative directions of stocks.

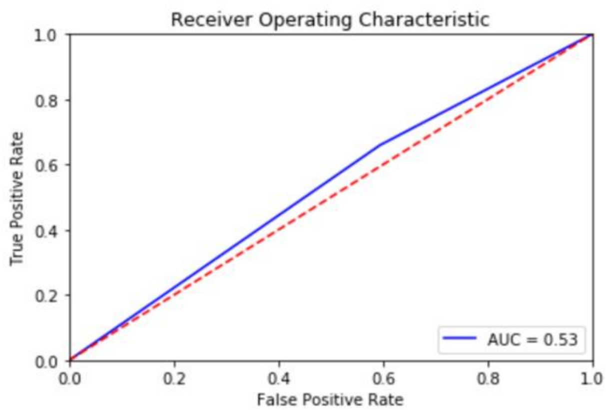


Figure 10. ROC curve of GARAN

4.2 Financial Metrics and Final Returns

In financial metrics part, first of all, we calculated the final cumulative returns for each stock and compared them with the benchmark strategy. Then, we examined five different financial metrics, CAGR, Annual Volatility, Maximum Drawdown, Sharpe and Calmar ratio, for both our model and the benchmark model, to evaluate the performance of the strategy in different manners.

Figure 11. Cumulative returns of GARAN and k-NN

The above graph shows the cumulative return of GARAN through years depending on our model and the benchmark model. (For all stocks, see Appendix). Our k-NN model has a positive return compared to buy and hold strategy for all k-values (see Appendix 7.1.6) for GARAN, yet this is not the case for other stocks. The below graph shows the case, for all stocks' final cumulative returns to see what will be the end equity in terms of percentage changes. Our model outperforms the benchmark model in 5 stocks which are AKBNK, EREGL, GARAN, TAHVL and TTKOM.

STOCK	K-NN	Buy & Hold
AKBNK	62.85%	31.91%
ARCLK	38.30%	229.11%
DOHOL	44.88%	96.23%
EREGL	1148.71%	482.79%
FROTO	287.05%	366.60%
GARAN	685.30%	56.39%
KCHOL	25.54%	194.75%
KOZAA	-92.62%	343.90%
SAHOL	-47.99%	55.72%
SISE	151.88%	255.86%
TAVHL	421.34%	289.07%
TCELL	-6.00%	85.76%
THYAO	383.84%	770.11%
TKFEN	-66.47%	400.23%
TTKOM	-21.33%	-23.32%

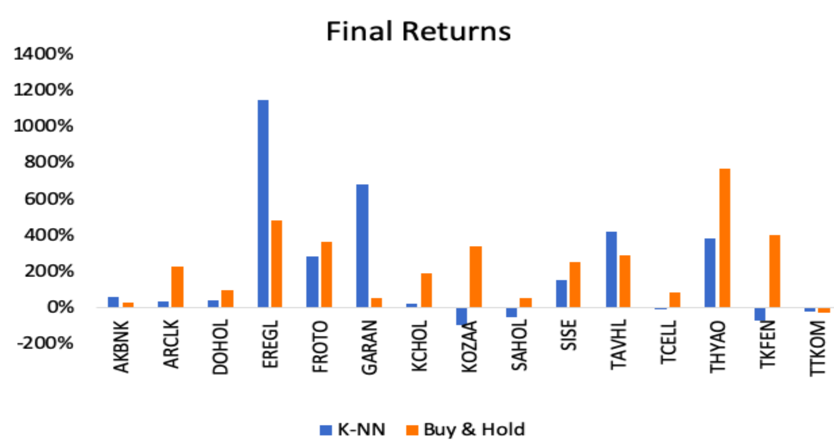


Figure 12. Final returns of k-NN and Buy & Hold strategy

In financial metrics part, we calculated each financial metric for each stock and strategy. *CAGR* gives an average annual return for the stock, *annual volatility* is a measure for risk that shows how volatile the return of the stock in a year, *maximum drawdown* shows the worst return of the stock in this period, *Sharpe ratio* gives risk-adjusted return of the stock, higher the ratio more the risk award of the stock, and the last metric *Calmar ratio*, is ratio of annual return of the stock to the maximum drawdown of it, and lower the ratio worse the performance of the stock in terms of risk-adjusted return.

	CAGR		Annual Volatility		Max Drawdown		Sharpe		Calmar	
STOCK	**K-NN**	**Buy & Hold**	**K-NN**	**Buy & Hold**	**K-NN**	**Buy & Hold**	**K-NN**	**Buy & Hold**	**K-NN**	**Buy & Hold**
AKBNK	0.07	-0.04	0.33	0.33	-0.43	-0.49	-0.38	-0.29	-0.17	-0.08
ARCLK	0.05	0.19	0.31	0.31	-0.58	-0.55	0.31	0.7	-0.08	0.34
DOHOL	0.05	0.1	0.43	0.43	-0.54	-0.63	0.34	0.44	0.1	0.16
EREGL	0.44	0.29	0.32	0.32	-0.29	-0.42	1.3	0.96	1.52	0.69
FROTO	0.22	0.25	0.32	0.32	-0.47	-0.42	0.77	0.85	0.46	0.59
GARAN	0.34	0.07	0.35	0.35	-0.31	-0.52	1.03	0.36	1.09	0.13
KCHOL	0.03	0.17	0.28	0.28	-0.59	-0.37	0.26	0.7	0.06	0.46
KOZAA	-0.31	0.24	0.59	0.59	-0.98	-0.89	-0.34	0.65	-0.32	0.26
SAHOL	-0.09	0.07	0.3	0.3	-0.78	-0.44	-0.16	0.36	-0.12	0.15
SISE	0.14	0.2	0.32	0.32	-0.46	-0.36	0.58	0.73	0.31	0.56
TAVHL	0.27	0.21	0.36	0.36	-0.41	-0.54	0.84	0.71	0.65	0.4
TCELL	-0.01	0.09	0.27	0.27	-0.7	-0.38	0.1	0.47	-0.01	0.25
THYAO	0.25	0.37	0.37	0.37	-0.47	-0.55	0.79	1.02	0.54	0.67
TKFEN	-0.15	0.26	0.35	0.35	-0.74	-0.53	-0.28	0.84	-0.2	0.48
TTKOM	-0.03	-0.04	0.29	0.29	-0.67	-0.56	0.02	0.01	-0.05	-0.07

Figure 13. Comparison of financial metrics

EREGL and GARAN have the highest CAGR in our model, with 0.44 and 0.34 which are quite high in financial markets. The benchmark model on average outperforms our model, average CAGR for it is around 16% which is 8% for us. Volatility is almost the same for both strategies returns, yet due to differences in returns and maximum drawdown, the k-NN model is riskier than the benchmark model.

Moreover, Maximum drawdown is better in the benchmark model; in our model we have even -0.98% which basically means loss of almost entire equity. Buy-and-hold model has on average -51% drawdown, and for our model, it is -57%. Sharpe ratio, as another parameter to risk and reward, is higher in k-NN model for 4 stocks, however, on average in our model Sharpe ratio is around 0.35 which is almost 0.6 for the benchmark model. The final metric, Calmar, gives an idea about the soundness of strategy, and our strategy outperforms the benchmark model, again, in 4 stocks, yet on average it is 0.25 for our strategy and 0.33 for the benchmark model. Taking all into account, we can say that, in terms of financial metrics, our model is not a solid strategy to implement in Turkish stock market.

5 Conclusion

In this paper, we used well-known k-NN classifier and investigated the feasibility of using it in Turkish stock market conditions, using historical data of 15 stocks out of 30 stocks of BIST30. We omitted transaction costs, and we used different technical indicators and metrics to evaluate our model. Yet, we could not find any significant result to claim that the k-NN model we built is feasible and applicable to the Turkish Stock Exchange Market. There is no specific k-value to apply to the whole stocks that we have investigated and our results with different k-values and different financial metrics do not give any strong path to follow in the market. We compared our model with the benchmark buy-and-hold and the benchmark model outperforms our model in terms of all five financial metrics and moreover, our classifier does not have the powerful capability of distinguishing positive or negative directions. For further research, other classifiers and machine learning models can be applied to the Turkish stock market such as Support Vector Machines(SVMs), Random Forest, and Neural Networks.

6 References

Aha, D., Kibler, D. and Albert, M. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1), pp.37-66.

Alkhatib, K., Najadat, H. and Hmeidi, I. (2013). Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm. *International Journal of Business, Humanities and Technology*, 3(3), pp.32-44.

Murphy, J. J. (1999). Technical analysis of the financial markets. New York Institute of Finance.

Teixeira, L. A., & De Oliveira, A. L. I. (2010). A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert systems with applications*, *37*(10), pp.6885-6890.

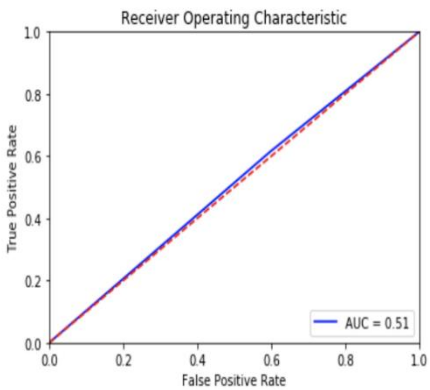
Theofilatos, K., Likothanassis, S., & Karathanasopoulos, A. (2012). Modeling and trading the EUR/USD exchange rate using machine learning techniques. *Engineering, Technology & Applied Science Research*, *2*(5), pp.269-272.

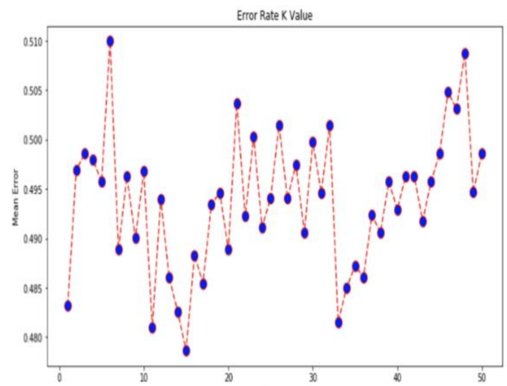
7 Appendix

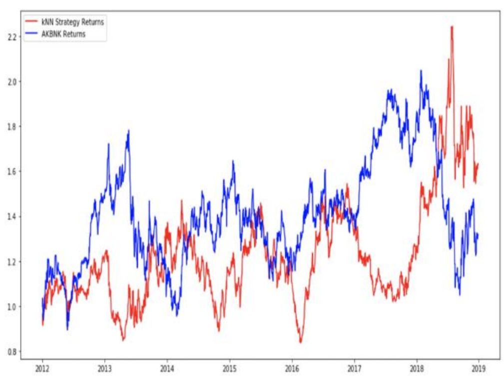
7.1 All Graphs

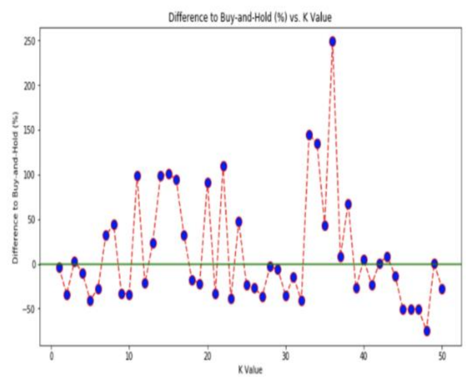
Graphs are reported respectively: ROC & AUC, error rate vs k-value, cumulative return, difference to buy-and-hold strategy vs k-value graphs for all stocks.

7.1.1 AKBNK

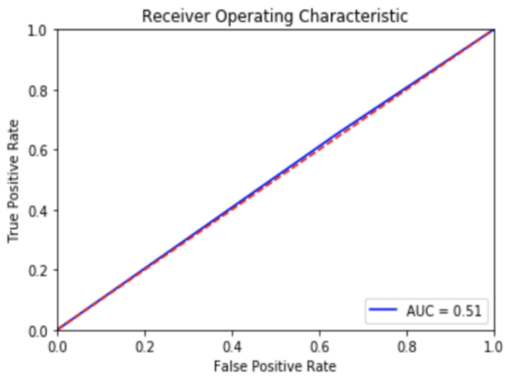


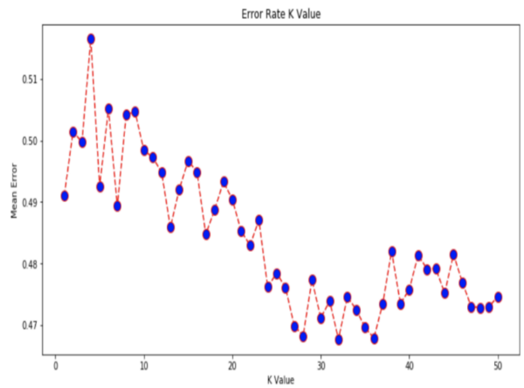


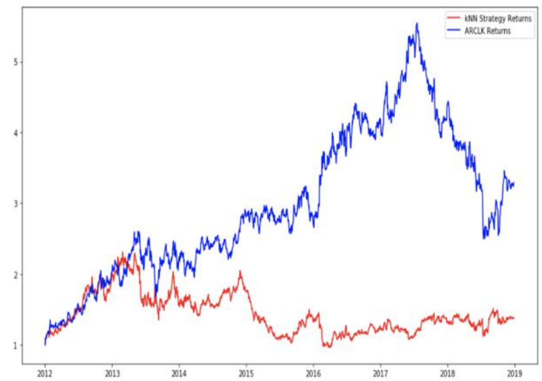


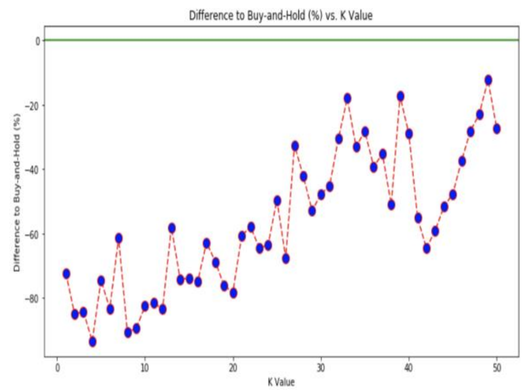


7.1.2 ARCLK

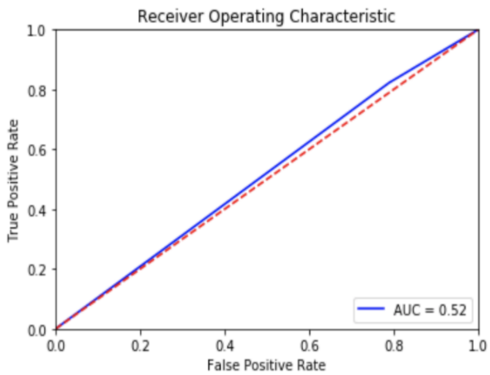

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.51
0.0
0.2
0.4
0.6
0.8
1.0

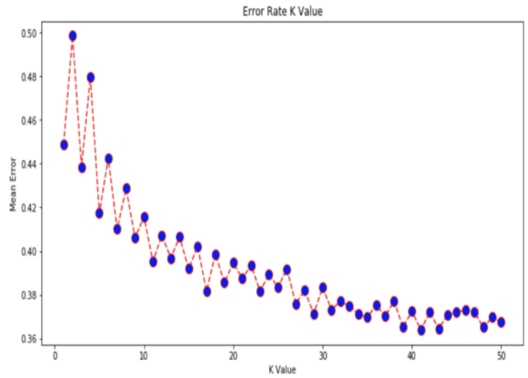

Error Rate K Value
Mean Error
K Value
0.47
0.48
0.49
0.50
0.51
0
10
20
30
40
50

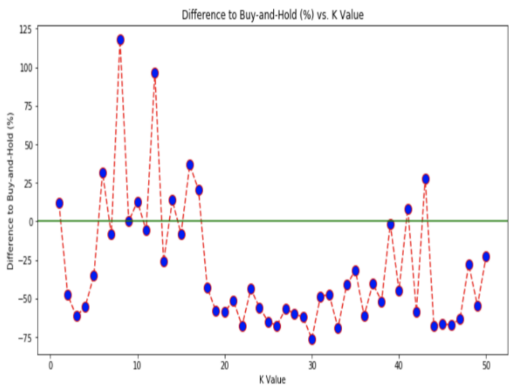

kNN Strategy Returns
ARCLK Returns
1
2
3
4
5
2012
2013
2014
2015
2016
2017
2018
2019

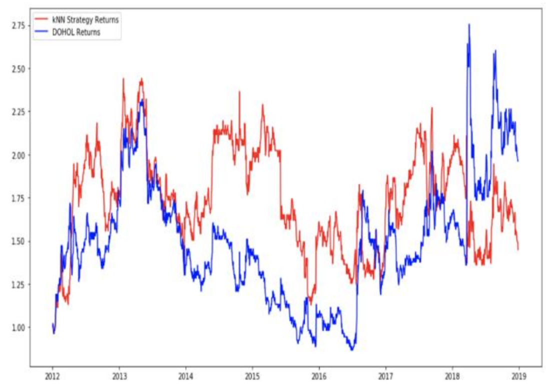

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value
0
-20
-40
-60
-80
0
10
20
30
40
50

7.1.3 DOHOL

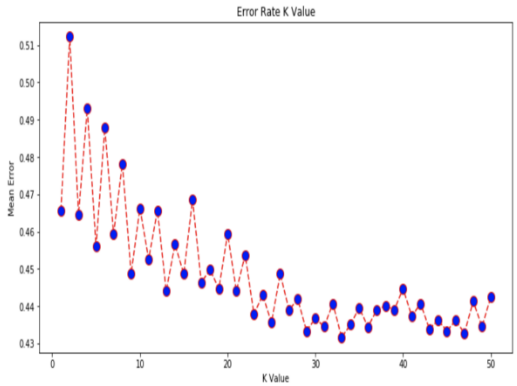

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.52

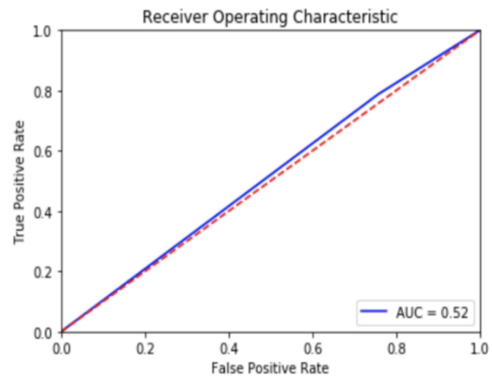

Error Rate K Value
Mean Error
K Value


Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

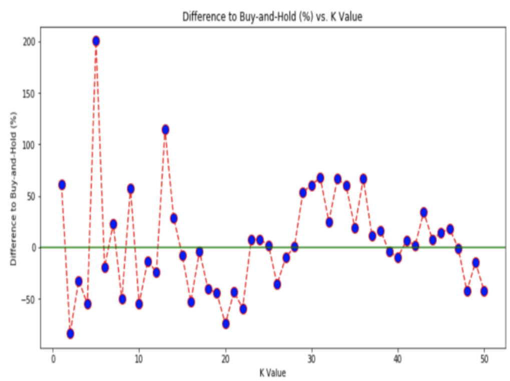

kNN Strategy Returns
DOHOL Returns

7.1.4 EREGL

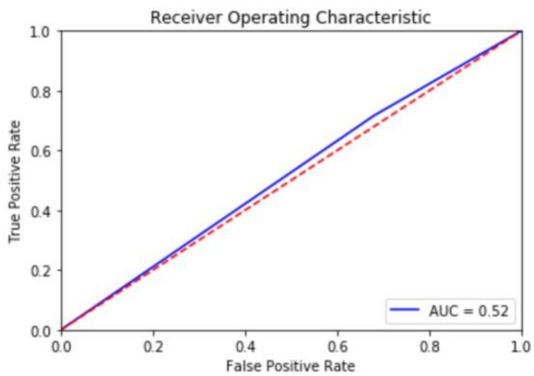

Error Rate K Value
Mean Error
K Value

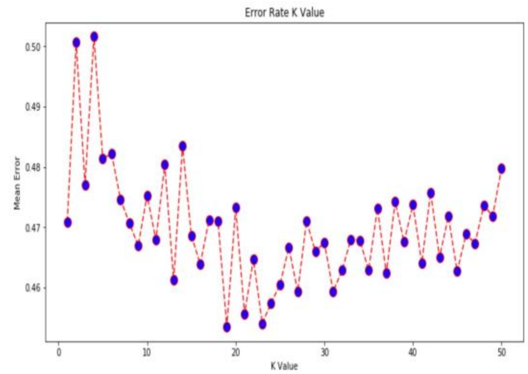

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.52

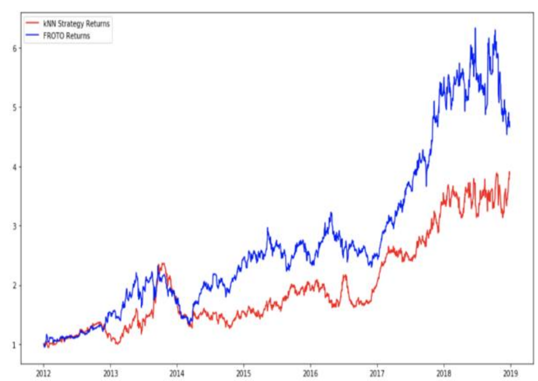

kNN Strategy Returns
EREGL Returns

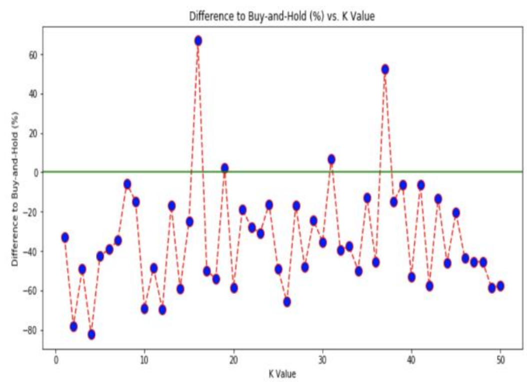

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.5 FROTO

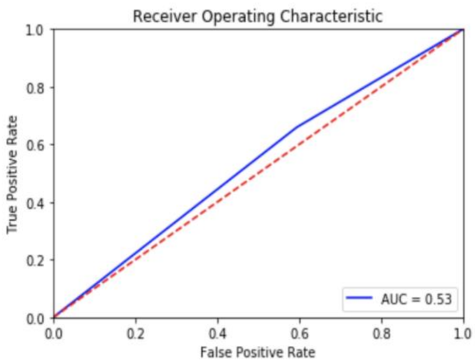

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.52

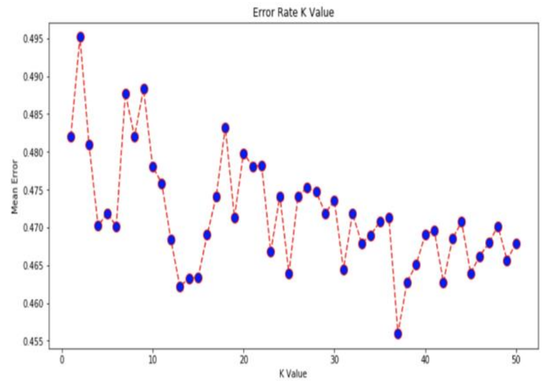

Error Rate K Value
Mean Error
K Value

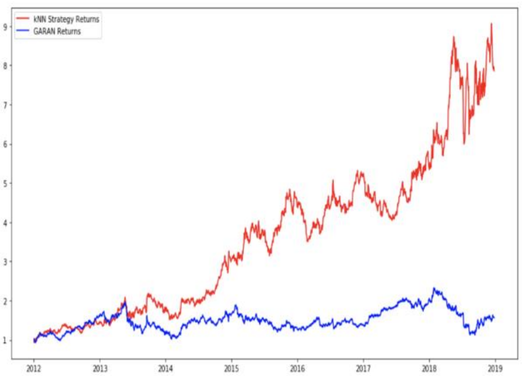

kNN Strategy Returns
FROTO Returns

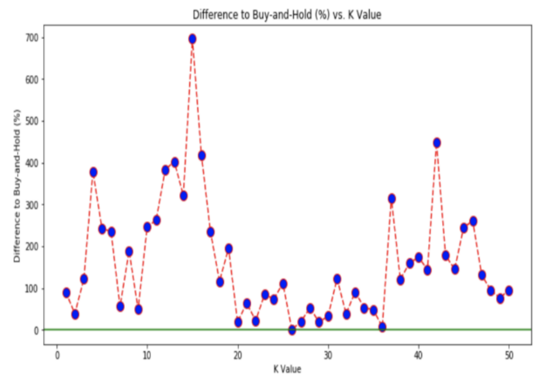

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.6 GARAN

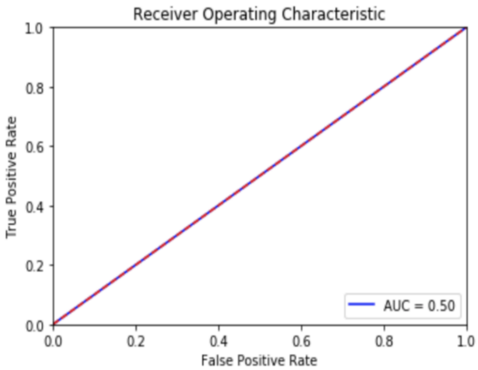

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.53

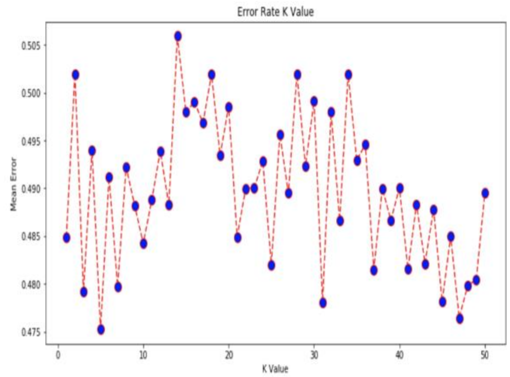

Error Rate K Value
Mean Error
K Value

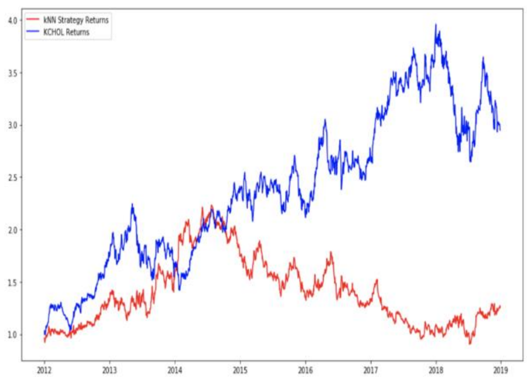

kNN Strategy Returns
GARAN Returns

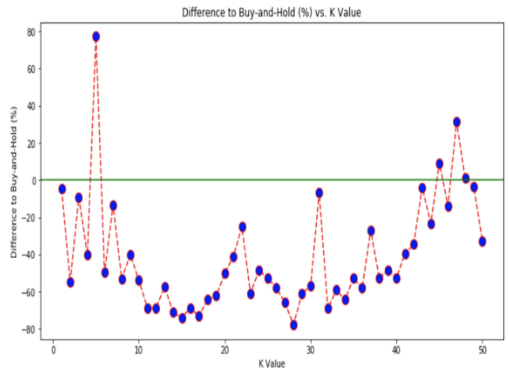

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.7 KCHO

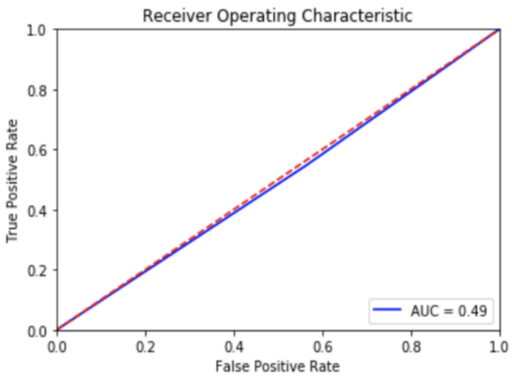

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.50

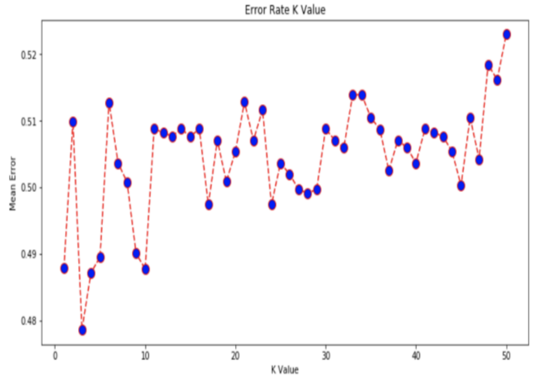

Error Rate K Value
Mean Error
K Value


kNN Strategy Returns
KCHOL Returns

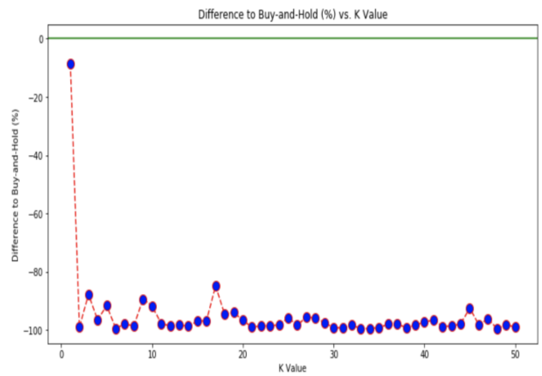

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.8 KOZAA

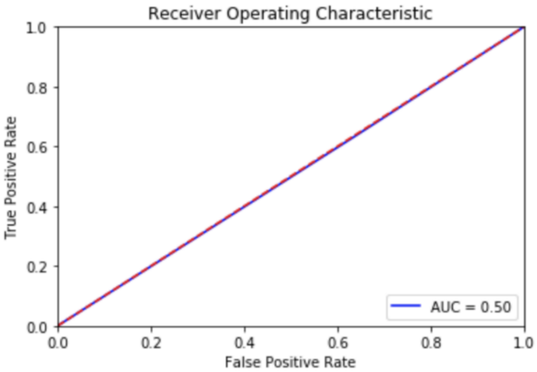

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.49
0.0
0.2
0.4
0.6
0.8
1.0

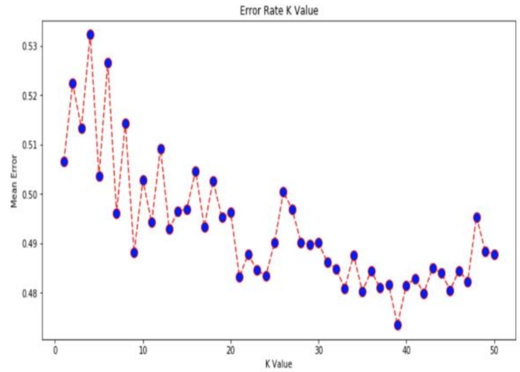

Error Rate K Value
Mean Error
K Value
0.48
0.49
0.50
0.51
0.52
0
10
20
30
40
50

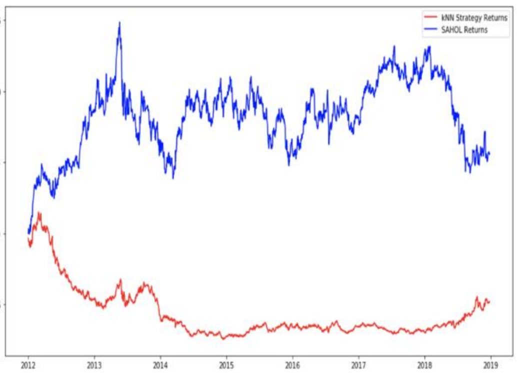

kNN Strategy Returns
KOZAA Returns
0
1
2
3
4
5
2012
2013
2014
2015
2016
2017
2018
2019

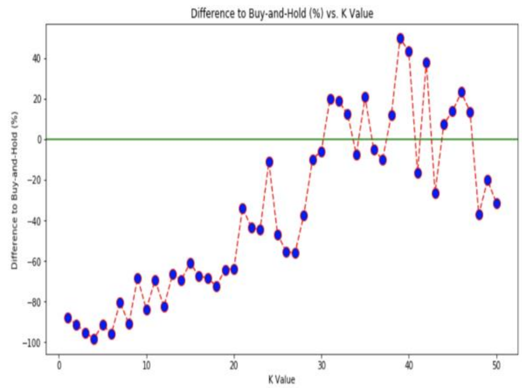

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value
0
−20
−40
−60
−80
−100
0
10
20
30
40
50

7.1.9 SAHOL

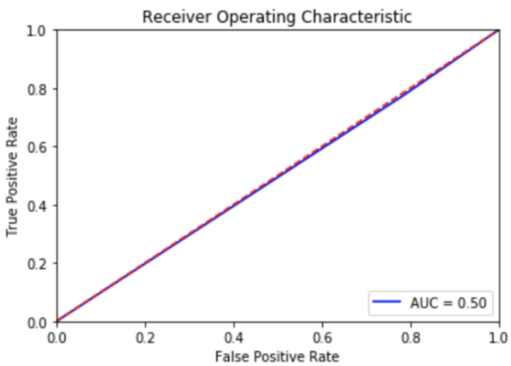

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.50

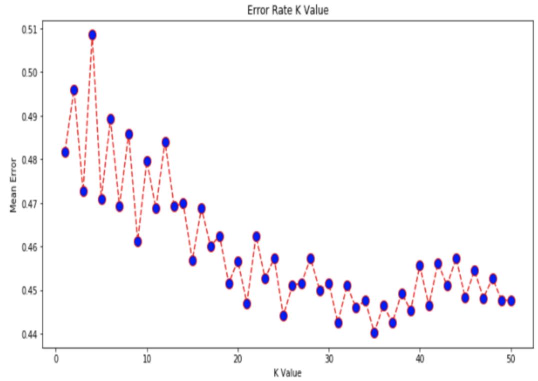

Error Rate K Value
Mean Error
K Value


kNN Strategy Returns
SAHOL Returns

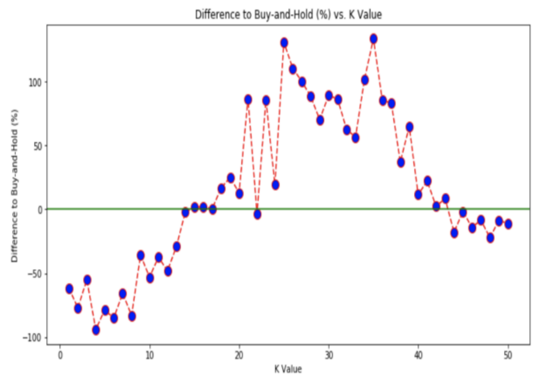

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.10 SISE

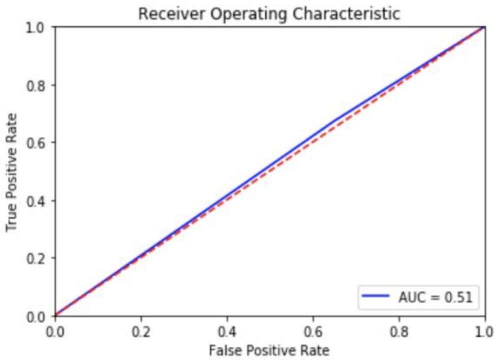

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.50

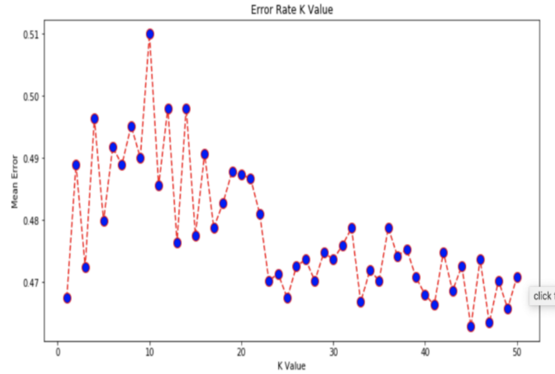

Error Rate K Value
Mean Error
K Value

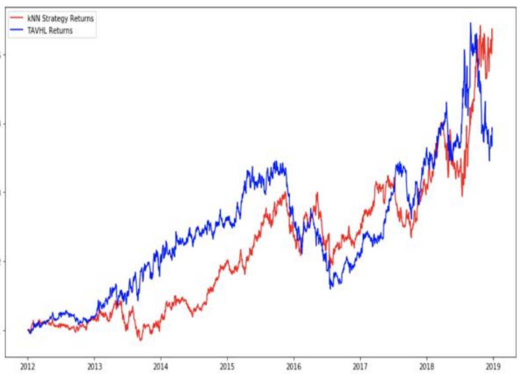

kNN Strategy Returns
SISE Returns

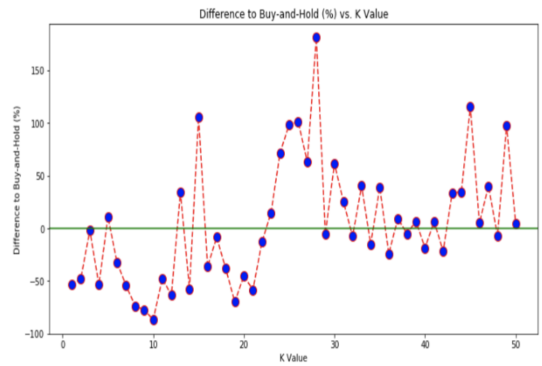

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.11 TAVHL

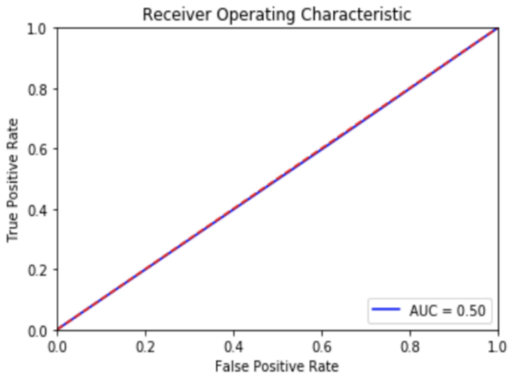

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.51

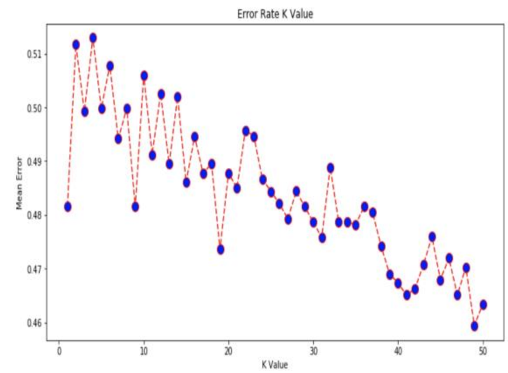

Error Rate K Value
Mean Error
K Value

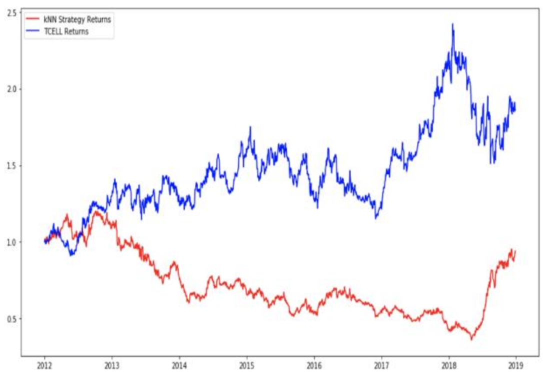

kNN Strategy Returns
TAVHL Returns

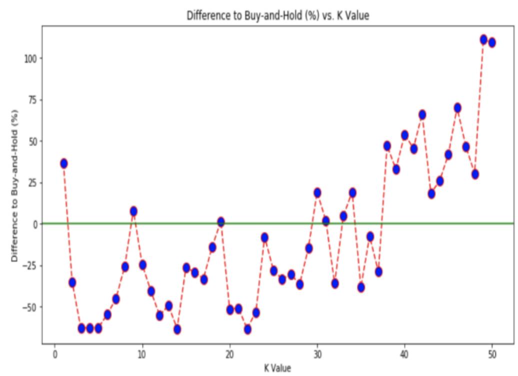

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.12 TCELL

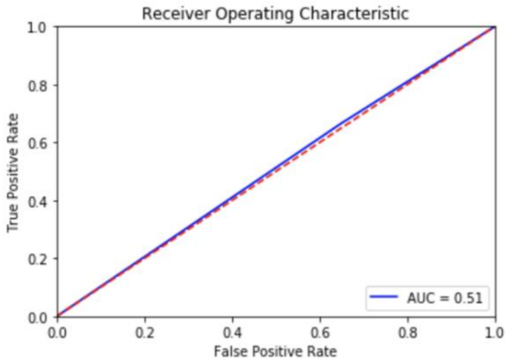

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.50

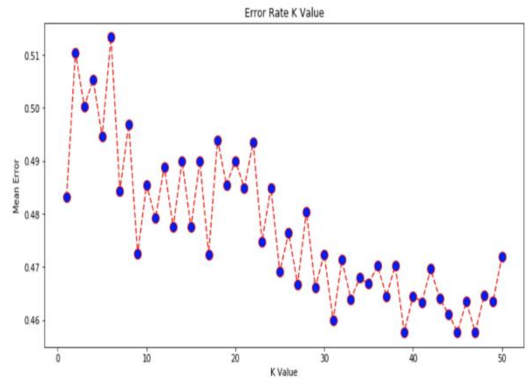

Error Rate K Value
Mean Error
K Value

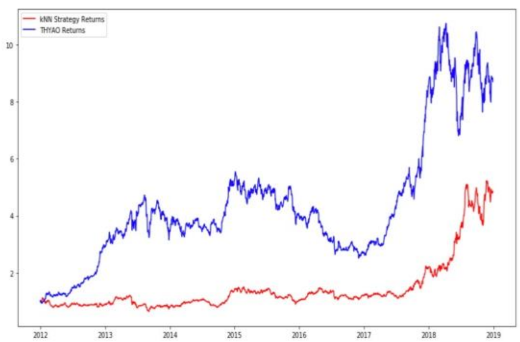

kNN Strategy Returns
TCELL Returns

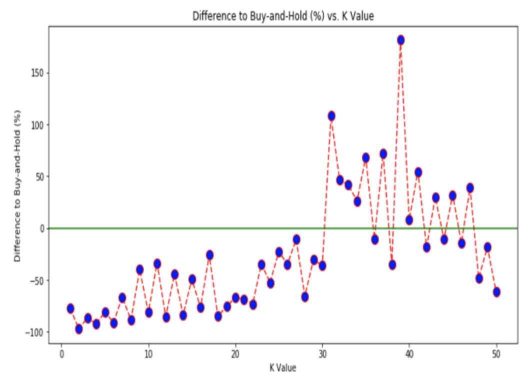

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.13 THYAO

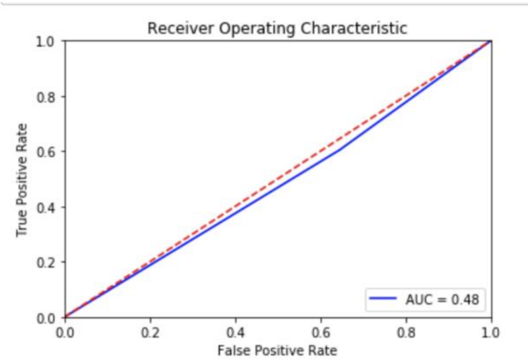

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.51

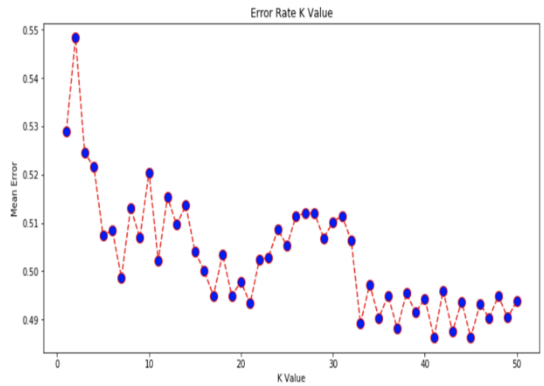

Error Rate K Value
Mean Error
K Value

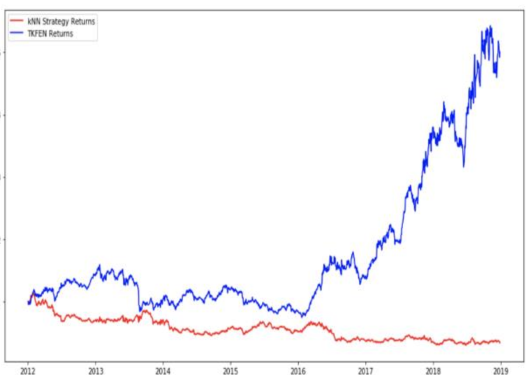

kNN Strategy Returns
THYAO Returns

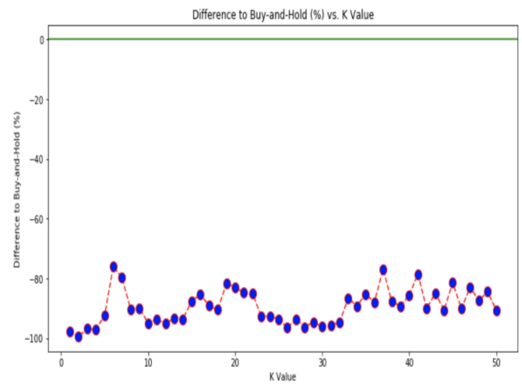

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.14 TKFEN

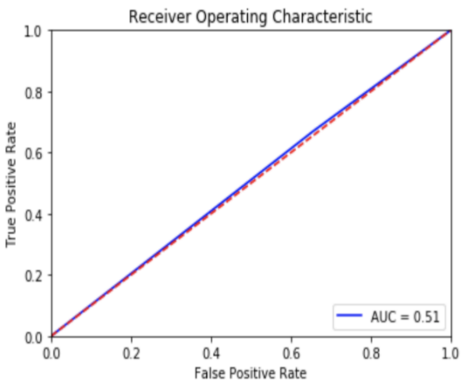

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.48
0.0
0.2
0.4
0.6
0.8
1.0

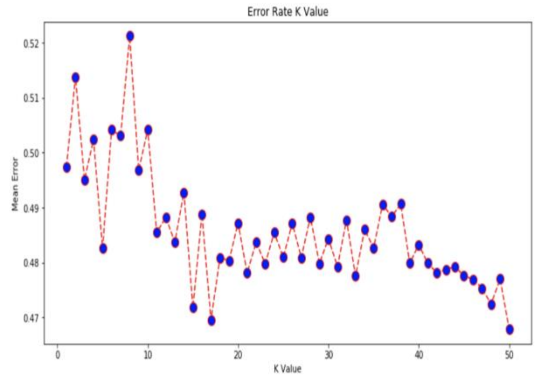

Error Rate K Value
Mean Error
K Value

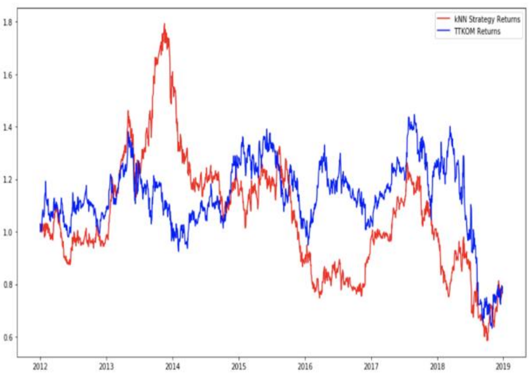

kNN Strategy Returns
TKFEN Returns
2012
2013
2014
2015
2016
2017
2018
2019

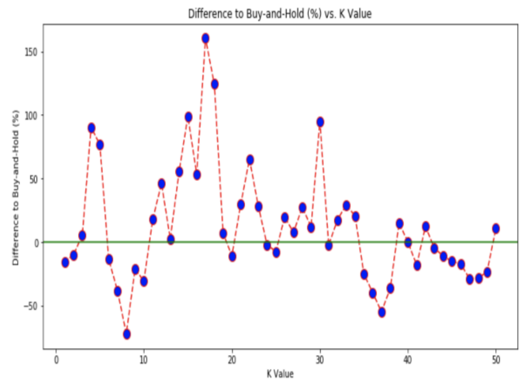

Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value

7.1.15 TTKOM


Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.51


Error Rate K Value
Mean Error
K Value


kNN Strategy Returns
TTKOM Returns


Difference to Buy-and-Hold (%) vs. K Value
Difference to Buy-and-Hold (%)
K Value