STOCK MARKET TREND PREDICTION USING K-NEAREST NEIGHBOR (KNN) ALGORITHM

[1] R.Sandhya Rani,[2] Mrs. G.Priyanka
[1]M.Tech Student,[2]Assistant Pofessor
Department Of Computer Science And Engineering
KLR College Of Engineering And Technology,Telangana

ABSTRACT

This paper examines a hybrid model which combines a K-Nearest Neighbors (KNN) approach with a probabilistic method for the prediction of stock price trends. One of the main problems of KNN classification is the assumptions implied by distance functions. The assumptions focus on the nearest neighbors which are at the centroid of data points for test instances. This approach excludes the non-centric data points which can be statistically significant in the problem of predicting the stock price trends. For this it is necessary to construct an enhanced model that integrates KNN with a probabilistic method which utilizes both centric and non-centric data points in the computations of probabilities for the target instances. The embedded probabilistic method is derived from Bayes' theorem. The prediction outcome is based on a joint probability where the likelihood of the event of the nearest neighbors and the event of prior probability occurring together and at the same point in time where they are calculated. The proposed hybrid KNN Probabilistic model was compared with the standard classifiers that include KNN, Naive Bayes, One Rule (OneR) andZero Rule (ZeroR). The test results showed that the proposed model outperformed the standard classifiers which were used for the comparisons.

Keywords: Stock Price Prediction, K-Nearest Neighbors, Bayes' Theorem, Naive Bayes, Probabilistic Method

I. INTRODUCTION

Analyzing financial data in securities has been an important and challenging issue in the investment community. Stock price efficiency for public listed firms is difficult to achievedue to the opposing effects of information competition among major investors and the adverse selection costs imposed by their information advantage.

There are two main schools of thought in analyzing the financial markets. The first approach is known as fundamental analysis. The methodology used in fundamental analysis evaluates a stock by measuring its intrinsic value through qualitative and quantitative analysis. This approach examines a company's financial reports, management, industry,micro and macro-economic factors.

The second approach is known as technical analysis. The methodology used in technical analysis for forecasting the direction of prices is through the study of historical market data. Technical analysis uses a variety of charts to anticipate what are likely to happen. The stock charts include candlestick charts, line charts, bar charts, point and figure charts, OHLC (open-high-low-close) charts and mountain charts. The charts are viewable in different time frames with price and volume. There are many types of indicators used in the charts, including resistance, support, breakout, trending and momentum.

Several alternatives to approach this type of problem have been proposed, which range from traditional statistical modelling to methods based on computational intelligence and machine learning. Vanstone and Tan surveyed the works in the domain of applying soft computing to financial trading and investment. They categorized the papers reviewed in the following areas: time series, optimization, hybrid methods, pattern recognition and classification. Within the context of financial trading discipline, the survey showed that most of the research was being conducted in the field of technical analysis. An integrated fundamental and technical analysis model was examined to evaluate the stock price trends by focusing on macro-economic analysis. It also analyzed the company behaviour and theassociated industry in relation to the economy which in turn provide more information for investors in their investment decisions.

A nearest neighbor search (NNS) method produced an intended result by the use of KNN technique with technical analysis. This model applied technical analysis on stockmarket data which include historical price and trading volume. It applied technical indicators made up of stop loss, stop gain and RSI filters. The KNN algorithm part applied the distance function on the collected data. This model was compared with the buy-and-hold strategy by using the fundamental analysis approach.

Fast Library for Approximate Nearest Neighbors (FLANN) is used to perform the searches for choosing the best algorithm found to work best among a collection of algorithms in its library. Majhi etal. examined the FLANN model to predict the S&P 500 indices, and the FLANN model was established by performing fast approximate nearest neighbor searchesin high dimensional spaces.

Artificial neural networks (ANN) exhibit high generalization power as compared to conventional statistical tools. ANN is able to infer from historical data to identify the characteristics of performing stocks. The information is reflected in technical and financial variables. As a result, ANN is used as a statistical tool to explore the intricate relationships between the related financial and technical variables and the performance of stocks.

Neural network modelling can decode nonlinear regularities in asset price movements. Statistical inference and modifications to standard learning techniques prove useful in dealing with the salient features of economic data.

Some research has been carried out through the use of both qualitative and quantitative analysis. Shynkevich et al. studied how the performance of a financial forecasting model was improved by the use of a concurrent, and appropriately weighted news articles, having different degrees of relevance to the target stock. The financial model supported the decision-making process of investors and traders. Textual pre-processing techniques were utilized for the predictive system. A multiple kernel learning technique was applied to predict the stock price movements. The technique integrated information extracted from multiple news categories while separate kernels were used to analyze each category. The news articles were partitioned according to some factors from the industries and their relevance to the target stock. The experiments were performed on stocks from the health care sector. The results showed that the financial forecasting model had achieved better performance when data sources contain increased categories of the number of relevant news. An enhanced model for this study incorporated additional data source using historical prices and made predictions based on both textual and time series data. Additional kernels can be employed for different data sources. The use of new categorical features was to improve the forecasting performance.

Linear regression is commonly used in financial analysis and forecasting. Many regression classifiers had demonstrated their usefulness to analyze quantitative data to make forecast by estimating the model parameters.

A regression driven fuzzy transform (RDFT) distributes a smoothing approximation of time series with a smaller delay as compared with moving average. This feature is important for forecasting tool where time plays a key role.

In high dimensional data, not all features are relevant and have an influence on the outputs. An Enhanced Feature Representation Based on Linear Regression Model for Stock Market Prediction was evaluated to investigate the statistical metrics used in feature selection that extracts the most relevant features to reduce the high dimensionality of the data. The statistical metrics include Information Gain, Term Frequency-Invert Document Frequency and the Document Frequency. The study illustrated that the identification of the relevant feature representations produced better result in the prediction output.

A hybrid intelligent data mining methodology based on Genetic Algorithm - Support Vector Machine Model was reviewed to explore stock market tendency. This approach makes use of the genetic algorithm for variable selection in order to improve the speed of support vector machine by reducing the model complexity, and then the historical data is used to identify stock market trends. Hybrid techniques can be used to improve the existing forecasting models due to the limitation of ANN like black box technique. A combination of methods such as fuzzy rule-based system, fuzzy neural network and Kalman filter with hybrid neuro-fuzzy architecture have been developed to predict financial time series data.

This research studies a hybrid approach through the use of KNN algorithm and a probabilistic method for predicting the stock price trends.

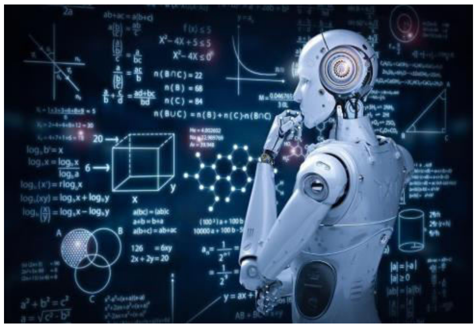
Comparison of various learning classifiers

When developing a classifier using various functions from different classifiers, it is important to compare the performances of the classifiers. Simulation results can provide us with direct comparison results for the classifiers with a statistical analysis of the objective functions. The hybrid KNN-Probabilistic model was compared with the supervised learning and classification algorithms, including ‘KNN’, ‘Naïve Bayes’, ‘OneR’ and ‘ZeroR’.

Machine Learning (ML):

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people. Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve.

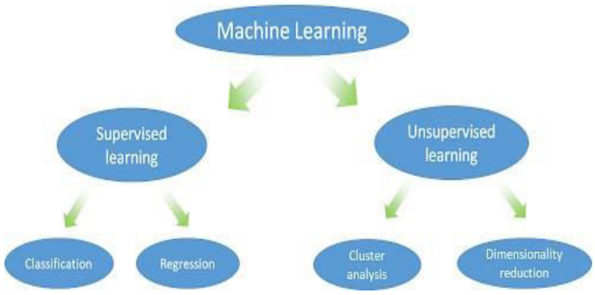
Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.



Machine Learning Methods:

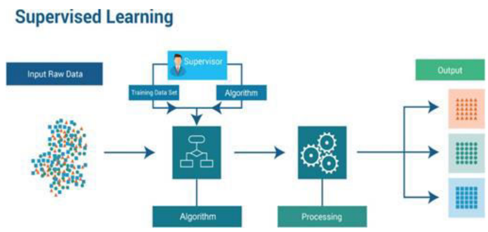
Two of the most widely adopted machine learning methods are

1. Supervised learning
2. Unsupervised learning



1. Supervised learning:

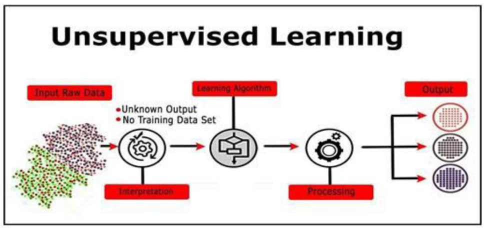
In supervised learning, the computer is provided with example inputs that are labeled with their desired outputs. The purpose of this method is for the algorithm to be able to “learn” by comparing its actual output with the “taught” outputs to find errors, and modify the model accordingly. Supervised learning therefore uses patterns to predict labelvalues on additional unlabeled data.



2. Unsupervised Learning:

In unsupervised learning, data is unlabeled, so the learning algorithm is left to find commonalities among its input data. As unlabeled data are more abundant than labeled data, machine learning methods that facilitate unsupervised learning are particularly valuable.

The goal of unsupervised learning may be as straightforward as discovering hidden patterns within a dataset, but it may also have a goal of feature learning, which allows the computational machine to automatically discover the representations that are needed to classify raw data. Unsupervised learning is commonly used for transactional data.



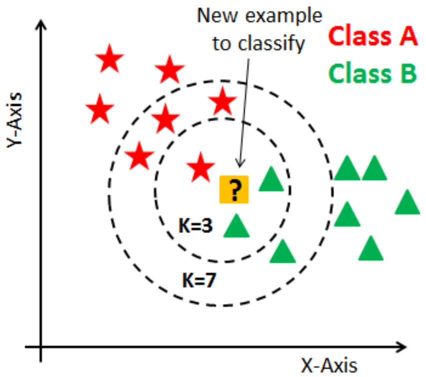
Without being told a “correct” answer, unsupervised learning methods can look at complex data that is more expansive and seemingly unrelated in order to organize it in potentially meaningful ways. Unsupervised learning is often used for anomaly detection including for fraudulent credit card purchases, and recommender systems that recommend what products to buy next.

Application of Machine Learning:

- Face Recognition
- Social Media Services
- Virtual Personal Assistants
- Online Fraud Detection
- Product Recommendations
- Autonomous
- Self-Driving Cars
- Stock Market Trading

K-Nearest Neighbors Classifier

The KNN algorithm is used to measure the distance between the given test instance and allthe instances in the data set, this is done by choosing the ‘k’ closest instances and then predict the class value based on these nearest neighbors. The ‘k’ is assigned as number of neighbors voting on the test instance. As such KNN is often referred to as case based learning or an instance-based learning where each training instance is a case from the problem domain. KNN is also referred to as a lazy learning algorithm due to the fact that there is no learning ofthe model required and all of the computation works happen at the time a prediction is requested. KNN is a non-parametric machine learning algorithm as it makes no assumptions about the functional form of the problem being solved. Each prediction is made for a new instance (x) by searching through the entire training set for the ‘k’ most nearest instances and applying majority voting rule to determine the prediction outcome.



Advantages of KNN Algorithm:

- It is simple to implement.
- It is robust to the noisy training data
- It can be more effective if the training data is large.

Disadvantages of KNN Algorithm:

- Always needs to determine the value of K which may be complex some time.
- The computation cost is high because of calculating the distance between the datapoints for all the training samples.

Bayes’ Theorem

The Bayes’ theorem plays an important role in probabilistic learning and classification. The Bayesian classification represents a supervised learning method as well as a statistical method for classification. It has learning and classification methods based on probability theory.

The Bayesian classification is named after Thomas Bayes, who proposed the Bayes’ theorem. Bayes’ theorem is often called Bayes’ rule. The Bayes’ rule uses prior probability of each category given no information about an item. Bayesian classification provides probabilistic methods where prior knowledge and observed data can be combined. It has a useful perspective for evaluating a variety of learning algorithms.

The Bayesian classification calculates explicit probabilities for hypothesis and it is also robust to noise in input data. Given a hypothesis h and data D which bears on the hypothesis, Bayes' theorem is stated as

$$P(h/D) = \frac{P(h/D)P(h)}{P(D)}$$

Where:

P(D): independent probability of D
P(h): independent probability of h: prior probability
P(h|D): conditional probability of h given D: posterior probability
P(D|h): conditional probability of D given h: likelihood

II. MATERIALS AND METHODS

Data were collected from Bursa Malaysia which is the main stock exchange in Malaysia. The data sources are corporate annual reports which include income statements, cash flow and balance sheet. The features in the data set were formulated based on fundamental analysis. The features were financial ratios which are indicators of the companies' financial health. Table 3 and Table 4 illustrate the structure of data set 1 and data set 2. Data set 1 contains original data from heterogeneous sources with different data types used for currency values and financial ratios. All of the variables in data set 1 have numerical data type. The data in data set 1 was transformed into data set 2 with variables in categorical data type. The transformed data set provides a way to measure rankings of stock price performance in a standardized data format. The transformation process is shown in Figure 1.

Figure 1, illustrates that initially the numerical data from corporate reports were used as inputs. At this stage the data are raw facts. The raw data are then used for calculations of the financial ratios. The values in financial ratios are fractional type. At this stage the data are semantic but diverse. After that, the fractional data are then converted into standardized percentage values based on performance interpretations of the features. At this stage the data are standardized. A ranking table is set up to group performance categories based on ranges of percentage values. The table is then used in data mapping to categorize the data from the percentage data format into the categorical data format. At this stage the data are interpretable.

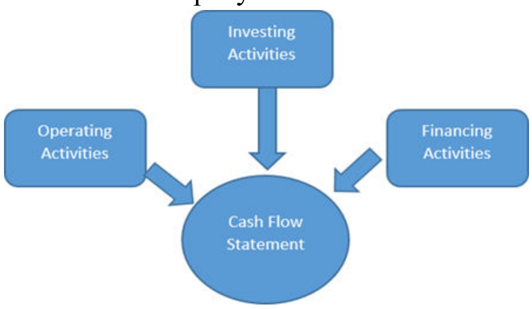
Debt_Equity (D/E):

The debt-to-equity (D/E) ratio is calculated by dividing a company's total liabilities by its shareholder equity. These numbers are available on the balance sheet of a company's financial statements. The debt-to-equity ratio measures a company's debt relative to the value of its net assets, a high debt/equity ratio is often associated with high risk; it means that a company has been aggressive in financing its growth with debt.

Debt/Equity= Total Liabilities/Total shareholder's Equity

Cash Flow:

Cash flow is a measure of a company's financial health in terms of incomings and outgoings of cash, representing the operating activities of a company.



Cash from Operating Activities:

The operating activities on the CFS include any sources and uses of cash from business activities. In other words, it reflects how much cash is generated from a company's products or services.

Cash Flow from Investing Activities:

Cash flow from investing activities is one of the sections on the cash flow statement that reports how much cash has been generated or spent from various investment-related activities in a specific period. Investing activities include purchases of physical assets, investments in securities, or the sale of securities or assets.

Negative cash flow is often indicative of a company's poor performance. However, negative cash flow from investing activities might be due to significant amounts of cash being invested in the long-term health of the company, such as research and development.

Before analysing the different types of positive and negative cash flows from investing activities, it's important to review where a company's investment activity falls within its financial statements.

Cash Flow from Financing Activities:

Cash flow from financing activities (CFF) is a section of a company's cash flow statement, which shows the net flows of cash that are used to fund the company. Financing activities include transactions involving debt, equity, and dividends.

Cash flow from financing activities provides investors with insight into a company's financial strength and how well a company's capital structure is managed.

Formula:

CFF = CED − (CD + RP)

Where:

CED = Cash in flows from issuing equity or debt

CD = Cash paid as dividends

RP = Repurchase of debt and equity

Return_on_Equity (ROE):

Return on equity is a measure of profitability based on how much profit a company generates with each dollar of stockholders' equity. ROE is considered a measure of how effectively management is using a company's assets to create profits.

Return on Equity=Net Income/Average Shareholders' Equity

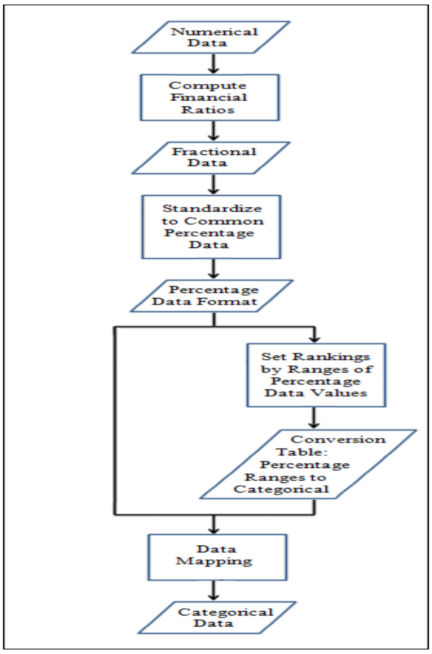


Figure 1: Flowchart of Data Transformation

Methods

The flow chart in Figure 2 illustrates the process flow of the proposed model. Using the parallel approach, the model starts with computing the prior probabilities and the probabilities based on KNN approach simultaneously on both the Profit class and Loss class. KNN initialization process involves the use of the k value for the nearest neighbors of test instances. KNN then calculates the number of Profit class and Loss class instances based on the k number of nearest neighbors in the vicinity of each test instance. The outcome generated from KNN is then used by the probabilistic method for further classification.

The probabilistic method calculates the prior probabilities of Profit class and Loss class based on the number of instances in the data set. The outcome from the earlier KNN approach is used as an input as the probabilities of Profit class and Loss class by the nearest neighbors method. The joint probabilities of Profit class and Loss class can then be calculated using the outcomes of the prior probabilities and the calculated KNN's probabilities. Finally, the predictive decision is made by comparing the joint probabilities of Profit class and Loss class.

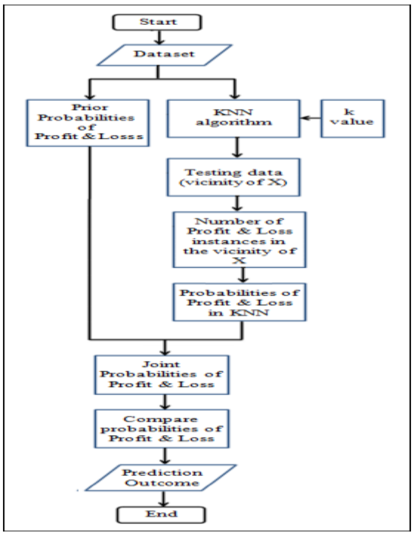


Figure 2: Flowchart of KNN-Probabilistic Model

III. RESULTS AND DISCUSSION

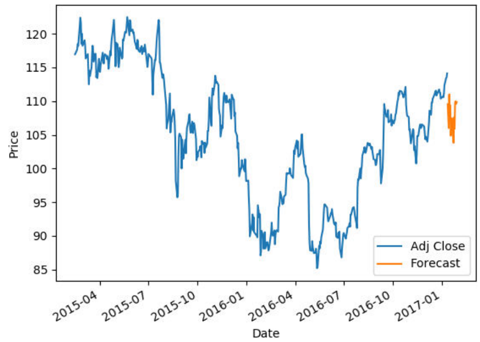
Results

The proposed model was tested and compared with four other standard algorithms, including KNN, Naïve Bayes, OneR and ZeroR. The test examined how accurate the tested algorithms predict the stock price trends, and evaluated the MAE and RMSE. Table 5 presents the test results. The hybrid KNN-Probabilistic model has allowed us to achieve an estimated accuracy of 89.1725%, exceeding the stand alone KNN reported accuracy of86.6667% and the Naive Bayes accuracy of 76.1194%. The accuracy rates for OneR and ZeroR classifiers were 71.6418% and 64.1791% respectively. KNN-Probabilistic model has MAE rate of 0.0667% and RMSE rate of 0.2582% which are much lower than the other classifiers.

Table 1: Prediction Results of Classifiers.

Classifier	Accuracy (%)	MAE	RMSE
KNN-Probabilistic	93.3333	0.0667	0.2582
KNN	86.6667	0.1333	0.3651
Naive Bayes	76.1194	0.1726	0.2824
OneR	71.6418	0.5325	0.6139
ZeroR	64.1791	0.4619	0.4805

Overall, KNN-Probabilistic model has better accuracy rate and error rates than the other classifiers used for comparisons. The test demonstrated that the hybrid mechanism of KNN and probabilistic method produced significantly improved results, compared with each of the KNN and Naïve Bayes classifiers.



Discussion

The proposed method begins with processing the data using data set 2, with each record contains a stock's financial features and the predicted outcomes in a structured categorical format. Using these records as inputs, stock price trends were predicted using the proposed hybrid KNN-Probabilistic model.

For KNN, the features in the data set are the data points in metric space with notion of distance. Each of the data set record contains a set of vectors and class labels associated with each vector. Each class label is either labeled as Profit for positive class or is labeled as Loss for negative class. The k value decides how many neighbors that can influence the classification. Initial step in KNN is to determine the appropriate k value. The k value is very training-data dependent. A small k value means that noise will have a higher influence on the result and a large value creates an overfit model. The use of k-fold cross-validation indicates the k value led to the highest classified generalizability. Typically, odd number is used as k value when the number of classes is two, so that a decision can be determined based on the class value with the higher number of instances.

For KNN-Probabilistic model, an odd number k value is not required because a decision for prediction is not made at the initial stage of KNN classifier. A 'k' value of even number is used to prevent unnecessary bias of unequal representations of the two classes at the stage of KNN method. A decision for prediction will be made based on the combined outcomes of the KNN method and the probabilistic method. The KNN method determines the class instances that form the initial probabilities from the nearest neighbors' perspective. The probabilistic method makes use of a combination of probabilities in its decision making. When the inputs for prior probability and probability based on KNN are available, the predictive model can calculate the joint probability and make prediction on the class outcome.

The probabilistic model includes some functional relations between the unknown parameters and the observed data to allow us to make predictions. The goal of this statistical analysis is to estimate the unknown parameters in the proposed model. Initial stage includes identifying the optimal value for the 'k' parameter. The computations used in the model include prior probability, probability in KNN and joint probability. The model has the following estimating computations.

The probability estimations based on KNN are,

- The probability of Profit = the number of Profit instances of the nearest neighbors in the vicinity of the query instance / Number of 'k' instances.
- The probability of Loss = the number of Loss instances of the nearest neighbors in the vicinity of the query instance / Number of 'k' instances.
- The probabilities in KNN measures the support provided by the data for each possible value of the 'k' parameter of KNN.

The computations of prior probability are,

- The prior probability of Profit = Total number of Profit instances / Total number ofall instances
- The prior probability of Loss = Total number of Loss instances / Total number of allinstances

The computations of joint probability are,

- Joint probability of Profit = the probability of Profit based on KNN × the priorprobability of Profit.
- Joint probability of Loss = the probability of Loss based on KNN × the priorprobability of Loss.

The steps of the process used for probability comparison is,

- If Joint Probability of Profit > Joint Probability of Loss Then prediction = Profit
- Else If Joint Probability of Loss > Joint Probability of Profit Then prediction = Loss
- Else If Joint Probability of Profit == Joint Probability of Loss Then repeat and re-adjust the k parameter.

IV. SCREEN SHOTS

Double click on 'run.bat' file to get below screen

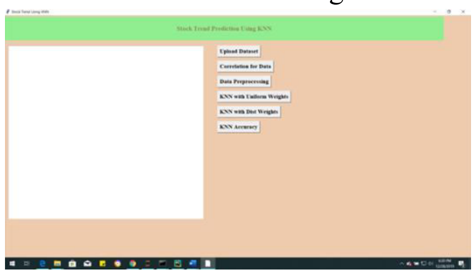


Figure 9.1 Stock Trend Prediction Using KNN

In above screen click on 'Download Button' download the Apple Stock and competitors data from Yahoo Finance Dataset

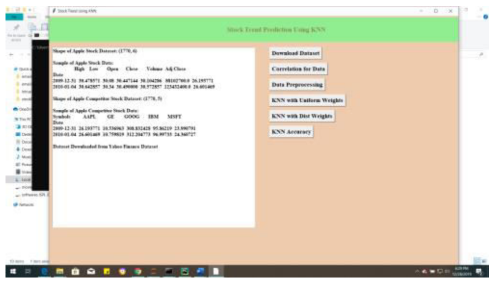


Figure 9.2 Download Dataset

In above screen I am Downloading of Apple Stock and Apple competitor Stock Data from Yahoo Finance Dataset.

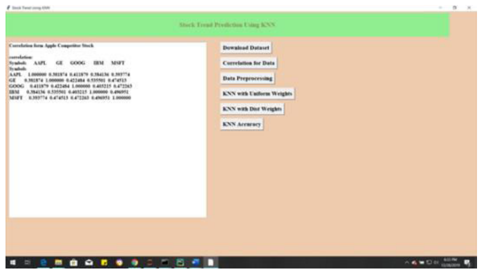


Figure 9.3 Correlations for Data

Now click on 'Correlation Data' Button to find the correlation between Apple and Competitor Stock market Dataset.

Show the trend in the technology industry rather than show how competing stocks affect each other.

Now click on 'Data Pre-processing button to drop missing values, split labels split train and test

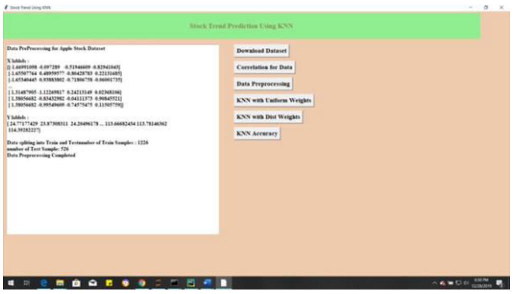


Figure 9.4 Data Pre-processing

After pre-processing all missing values are dropped, Separating the label here, Scalling of X, find Data Series of late X and early X (train) for model generation and evaluation, Separate label and identify it as y and Separation of training and testing of model.

In above screen we can see dataset contains total 1752 records and 1226 used for training and 526 used for testing.

Now click on 'Run KNN with Uniform Weights' to generate KNN model with uniform weights and calculate its model accuracy

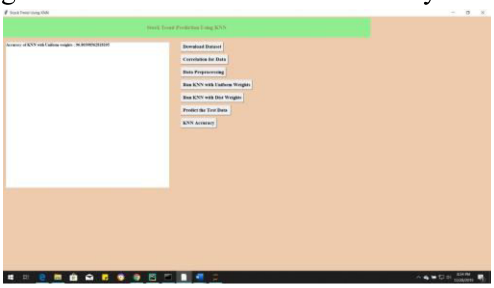


Figure 9.5 Run KNN with Uniform weight

In above screen we can see with KNN with uniform weights got 96.8% accuracy, now click on 'Run KNN with distance weights' to calculate accuracy

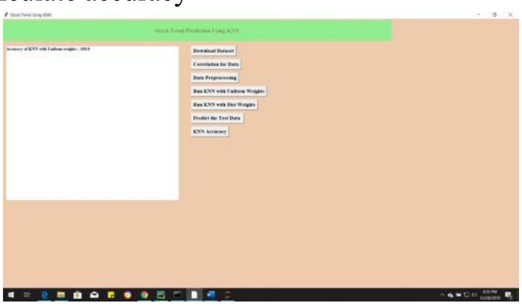


Figure 9.6 Run KNN with Dist. Weight

In above screen we got 100% accuracy, now we will click on 'Predict Test Data' button to upload test data and to predict whether test data stock market for both models.

Accuracy score (>0.95) for most of the models.

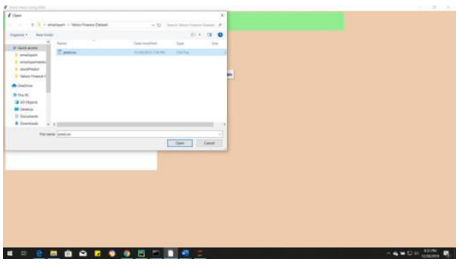


Figure 9.7 Test Data Upload

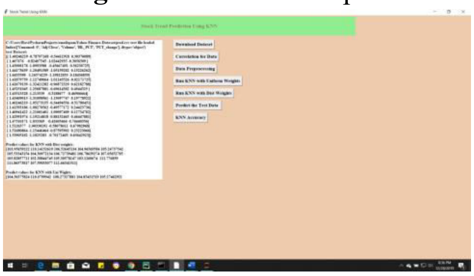


Figure 9.8 KNN Accuracy

In above screen for each test data we got forecast values for Apple Stock for each test record. Now click on 'KNN Accuracy' button to save the predicted values for each model save in the local directory and Accuracy comparison to both the models

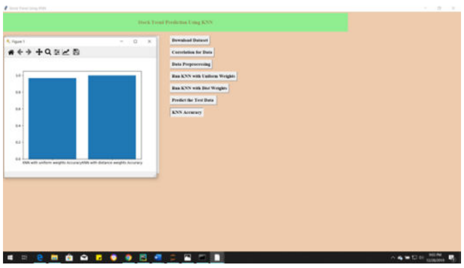
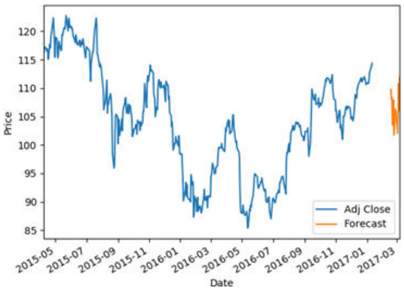


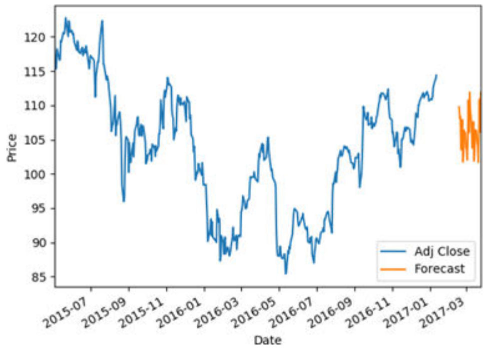
Figure 9.9 Test Data Results

From above graph we can see that distance weights has little bit better better accuracy compare to Uniform weights, in above graph x-axis contains algorithm name and y-axis represents accuracy of that algorithms

Plotting the Prediction for KNN with Uniform Weights:



Plotting the Prediction for KNN with Distance Weights:



V. CONCLUSION

The aim of this research is to improve the statistical fitness of the proposed model to overcome a KNN problem due to its computation approach. The KNN classifier can compute the empirical distribution over the Profit and Loss class values in the k number of nearest neighbors. However, the outcome is less than adequate due to sparse data. The KNN classifier has under fitting issue as it does not cater to generalization of sparse data outsidethe range of nearest neighborhood.

We have compared a hybrid KNN-Probabilistic model with four standard algorithms on the problem of predicting the stock price trends. Our results showed that the proposed KNN-Probabilistic model leads to significantly better results compared to the standard KNN algorithm and the other classification algorithms.

The limitation of the proposed model is that it applies a binary classification technique. The actual output of this binary classification model is a prediction score in two- class. The score indicates the model's certainty that the given observation belongs to either the Profit class or Loss class. For future work, the knowledge component is to transform the binary classification into multiclass classification. The multiclass classification involves observation and analysis of more than the existing two statistical class values. Additional research will include the application of the probabilistic model to multiclass data in order to provide more specific information of each class value. The newly formed multiclass classification will contain five class labels named “Sell”, “Underperform”, “Hold”, “Outperform”, and “Buy”. In numerical values for mapping purpose, we will convert “Sell”to -2 which implies strongly unfavorable; “Underperform” to -1 which implies moderately unfavorable; “Hold” to 0 which implies neutral; “Outperform” to 1 which implies moderately favorable; and “Buy” to 2 which implies strongly favorable.

BIBLIOGRAPHY

REFERENCES

[1] Benjamin Graham, Jason Zweig, and Warren E. Buffett, The Intelligent Investor, Publisher: Harper Collins Publishers Inc, 2003.

[2] Charles D. Kirkpatrick II and Julie R. Dahlquist, Technical Analysis: The Complete Resource for Financial Market Technicians (3rd Edition), Pearson Education, Inc., 2015.

[3] Bruce Vanstone and Clarence Tan, A Survey of the Application of Soft Computing to Investment and Financial Trading, Proceedings of the Australian and New Zealand Intelligent Information Systems Conference, Vol.

[4] Monica Tirea and Viorel Negru, Intelligent Stock Market Analysis System - A Fundamental and Macro-economical Analysis Approach, IEEE, 2014.

[5] Kian-Ping Lim, Chee-Wooi Hooy, Kwok-Boon Chang, and Robert Brooks, Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity, The North American Journal of Economics and Finance. Vol. 36, http://linkinghub.elsevier.com/retrieve/pii/S1 06 2940815001230, 2016, pp. 1–28.

[6] Lamartine Almeida Teixeira and Adriano Lorena Inácio de Oliveira, A method for automatic stock trading

combining technical analysis and nearest neighbor classification, Expert Systems with Applications, http://linkinghub.elsevier.com/retrieve/pii/S0 95 7417410002149, 2010, pp. 6885–6890.
[7] Banshidhar Majhi, Hasan Shalabi, and Mowafak Fathi, FLANN Based Forecasting of S&P 500 Index. Information Technology Journal, Vol. 4, Issue 3, http://www.scialert.net/abstract/?doi=itj.200 5.2 89.292, 2005, pp. 289–292.
[8] Ritanjali Majhi, G. Panda, and G. Sahoo, Development and performance evaluation of FLANN based model for forecasting of stock markets, Expert Systems with Applications, Vol. 36, Issue 3, http://linkinghub.elsevier.com/retrieve/pii/S0 95 7417408005526, 2009, pp. 6800–6808.
[9] Tong-Seng Quah and Bobby Srinivasan, Improving returns on stock investment through neural network selection, Expert Systems with Applications, Vol. 17, Issue 4, http://linkinghub.elsevier.com/retrieve/pii/S0 95 741749900041X, 1999, pp. 295–301.
[10] Halbert White, Economic prediction using neural networks: the case of IBM daily stock returns, IEEE International Conference on Neural Networks, http://ieeexplore.ieee.org/document/23959/, IEEE, 1988, pp. 451–458.
[11] Yauheniya Shynkevicha, T.M. McGinnity, Sonya A. Coleman, and Ammar Belatreche, Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning, Decision Support Systems, Vol. 85, http://linkinghub.elsevier.com/retrieve/pii/S0 16 7923616300252, 2016, pp. 74–83.
[12] Han Lock Siew and Md Jan Nordin, Regression Techniques for the Prediction of Stock Price Trend, 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), IEEE, 2012.
[13] Chi Ma, Junnan Liu, Hongyan Sun, and Haibin Jin, A hybrid financial time series model based on neural networks, IEEE, 2017.
[14] Lean Yu, Shouyang Wang, and Kin Keung Lai, Mining Stock Market Tendency Using GABased Support Vector Machines, Internet and Network Economics, http://link.springer.com/10.1007/11600930_ 33, 2005, pp. 336–345.
[15] Fu-Yuan Huang, Integration of an Improved Particle Swarm Algorithm and Fuzzy Neural Network for Shanghai Stock Market Prediction, 2008 Workshop on Power Electronics and Intelligent Transportation System.[Online].August 2008, IEEE, http://ieeexplore.ieee.org/document/4634852 /, 2008, pp. 242–247.
[16] Ching-hsue Cheng, Tai-liang Chen, and Hiajong Teoh, Multiple-Period Modified Fuzzy Time-Series for Forecasting TAIEX, Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), http://ieeexplore.ieee.org/document/4406190 /, IEEE, 2007, pp. 2–6.
[17] Pei-Chann Chang, Chin-Yuan Fan, and ShihHsin Chen, Financial Time Series Data Forecasting by Wavelet and TSK Fuzzy Rule Based System, Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), http://ieeexplore.ieee.org/document/4406255 /, IEEE, 2007, pp. 331–335.
[18] Lixin Yu and Yan-Qing Zhang, Evolutionary Fuzzy Neural Networks for Hybrid Financial Prediction, IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), Vol. 35, Issue 2, http://ieeexplore.ieee.org/document/1424198 /, IEEE, 2005, pp. 244–249.
[19] Chokri Slim, Neuro-fuzzy network based on extended Kalman filtering for financial time series, World Academy of Science, Engineering and Technology, Vol. 15, http://citeseerx.ist.psu.edu/viewdoc/downloa d?d oi=10.1.1.193.4464&rep=rep1&type=pdf,

2006, pp. 134–139.

[20] Gérard Biau and Luc Devroye, Lectures on the Nearest Neighbor Method, Springer, 2015.

[21] Saed Sayad, K Nearest Neighbors - Classification, [Online]. 2017, http://www.saedsayad.com/k_nearest_neighbors .htm, 2017, [Accessed: 10 January 2017].

[22] Dan Morris, Bayes' Theorem: A Visual Introduction for Beginners, Blue Windmill Media, 2017.

[23] James V Stone, Bayes' Rule With R: A Tutorial Introduction to Bayesian Analysis, Sebtel Press, 2016.

[24] Marco Scutari and Jean-Baptiste Denis, Bayesian Networks: With Examples in R, CRC Press, 2014.

[25] Peter Bruce and Andrew Bruce, Practical Statistics for Data Scientists: 50 Essential Concepts, O'Reilly Media, 2017.

[26] Ciaran Walsh, Key Management Ratios, 4th Edition (Financial Times Series), Prentice Hall, 2009.

[27] Seyed Enayatolah Alavi, Hasanali Sinaei, and Elham Afsharirad, Predict the Trend of Stock Prices Using Machine Learning Techniques, IAIEST, 2015.

[28] Lock Siew Han and Md Jan Nordin, Integrated Multiple Linear Regression-One Rule Classification Model for the Prediction of Stock Price Trend, Journal of Computer Sciences, Vol 13 (9), 2017, pp. 422-429.

[29] Abdolhossein Zameni and Othman Yong, Share Price Performance of Malaysian IPOs Around Lock-Up Expirations, Advanced Science Letters, Vol 23(9), 2017, pp. 8094-8102.

[30] Abdolhossein Zameni and Othman Yong, LockUp Expiry And Trading Volume Behaviour Of Malaysian Initial Public Offering's, International Journal of Economics and Financial Issues, Vol 6(3), 2016, pp. 12-21.

[31] Hani A.K. Ihlayyel, Nurfadhlina Mohd Sharef, Mohd Zakree Ahmed Nazri, and Azuraliza Abu Bakar, An Enhanced Feature Representation Based On Linear Regression Model For Stock Market Prediction, Intelligent Data Analysis, Vol 22(1), 2018, pp. 45-76.

[32] Lida Nikmanesh and Abu Hassan Shaari Mohd Nor, Macroeconomic Determinants of Stock Market Volatility: An Empirical Study of Malaysia and Indonesia, Asian Academy of Management Journal, Vol 21(1), 2016, pp. 161- 180.

[33] Luigi Troiano, Pravesh Kriplani, and Irene Díaz, Regression Driven F-Transform and Application to Smoothing of Financial Time Series, IEEE, 2017.