**LITERATURE SURVEY**

# 1) Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement

**AUTHORS:** [**Ernest Kwame Ampomah**](https://sciprofiles.com/profile/1101750)**,** [**Zhiguang Qin**](https://sciprofiles.com/profile/635950)

Forecasting the direction and trend of stock price is an important task which helps investors to make prudent financial decisions in the stock market. Investment in the stock market has a big risk associated with it. Minimizing prediction error reduces the investment risk. Machine learning (ML) models typically perform better than statistical and econometric models. Also, ensemble ML models have been shown in the literature to be able to produce superior performance than single ML models. In this work, we compare the effectiveness of tree-based ensemble ML models (Random Forest (RF), XGBoost Classifier (XG), Bagging Classifier (BC), AdaBoost Classifier (Ada), Extra Trees Classifier (ET), and Voting Classifier (VC)) in forecasting the direction of stock price movement. Eight different stock data from three stock exchanges (NYSE, NASDAQ, and NSE) are randomly collected and used for the study. Each data set is split into training and test set. Ten-fold cross validation accuracy is used to evaluate the ML models on the training set. In addition, the ML models are evaluated on the test set using accuracy, precision, recall, F1-score, specificity, and area under receiver operating characteristics curve (AUC-ROC). Kendall W test of concordance is used to rank the performance of the tree-based ML algorithms. For the training set, the AdaBoost model performed better than the rest of the models. For the test set, accuracy, precision, F1-score, and AUC metrics generated results significant to rank the models, and the Extra Trees classifier outperformed the other models in all the rankings.

# 2) Machine Learning versus Economic Restrictions: Evidence from Stock Return Predictability

## AUTHORS: [Doron Avramov](https://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=265949), [Si Cheng](https://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=1265857)

This paper shows that investments based on deep learning signals extract profitability from difficult-to-arbitrage stocks and during high limits-to-arbitrage market states. In particular, excluding microcaps, distressed stocks, or episodes of high market volatility considerably attenuates profitability. Machine learning-based performance further deteriorates in the presence of reasonable trading costs due to high turnover and extreme positions in the tangency portfolio implied by the pricing kernel. Despite their opaque nature, machine learning methods successfully identify mispriced stocks consistent with most anomalies. Beyond economic restrictions, deep learning signals are profitable in long positions and recent years and command low downside risk.

**3) A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction**

**AUTHORS** **:** [**Fitriana Harahap**](https://ieeexplore.ieee.org/author/37086246130)**;**[**Ahir Yugo Nugroho Harahap**](https://ieeexplore.ieee.org/author/37086805566)**;**[**Evri Ekadiansyah**](https://ieeexplore.ieee.org/author/37086806675)

# To choose the right vehicle according to the needs and funds owned by consumers, requires a careful analysis that takes into account many criteria and factors. The criteria used as a benchmark in choosing a vehicle, among others, price, spare parts, cylinder volume, the power of the vehicle. To process all these criteria required a system that can select and classify criteria chosen by consumer, so that can assist consumer in choosing the most appropriate vehicle, therefore needed a system for decision making in making car purchase. The Naive Bayes algorithm is a simple probabilistic classifier that computes a set of probabilities by summing the frequency and value combinations of the given dataset. Application of Naïve Bayes method is expected to be able to predict car purchases. Of the 20 car purchase data used in the test by the Naïve Bayes method, then obtained a percentage of 75% for the accuracy of prediction, where from 20 car purchase data tested there are 15 data purchase car successfully classified correctly.

# 4) Data classification using Support vector Machine (SVM), a simplified approach

**AUTHORS :** **[YingjunChen](https://www.sciencedirect.com/science/article/abs/pii/S0957417417301367" \l "!)****[YongtaoHao](https://www.sciencedirect.com/science/article/abs/pii/S0957417417301367" \l "!)**

This study investigates stock market indices prediction that is an interesting and important research in the areas of investment and applications, as it can get more profits and returns at lower risk rate with effective exchange strategies. To realize accurate prediction, various methods have been tried, among which the [machine learning methods](https://www.sciencedirect.com/topics/engineering/machine-learning-method) have drawn attention and been developed. In this paper, we propose a basic hybridized framework of the feature weighted [support vector machine](https://www.sciencedirect.com/topics/engineering/support-vector-machine) as well as feature weighted K-nearest neighbor to effectively predict stock market indices. We first establish a detailed theory of feature weighted SVM for the [data classification](https://www.sciencedirect.com/topics/computer-science/data-classification) assigning different weights for different features with respect to the classification importance. Then, to get the weights, we estimate the importance of each feature by computing the information gain. Lastly, we use feature weighted K-nearest neighbor to predict future stock market indices by computing k weighted nearest neighbors from the historical dataset. Experiment results on two well known Chinese stock market indices like Shanghai and Shenzhen stock exchange indices are finally presented to test the performance of our established model. With our proposed model, it can achieve a better prediction capability to Shanghai Stock Exchange Composite Index and Shenzhen Stock Exchange Component Index in the short, medium and long term respectively. The proposed algorithm can also be adapted to other stock market indices prediction.

# 5) Artificial Neural Networks architectures for stock price prediction: comparisons and applications

# AUTHORS: Luca Di Persio, Luca Di Persio

We present an Artificial Neural Network (ANN) approach to predict stock market indices, particularly with respect to the forecast of their trend movements up or down. Exploiting different Neural Networks architectures, we provide numerical analysis of concrete financial time series. In particular, after a brief resum ´ e of the ´ existing literature on the subject, we consider the Multi-layer Perceptron (MLP), the Convolutional Neural Networks (CNN), and the Long Short-Term Memory (LSTM) recurrent neural networks techniques. We focus on the importance of choosing the correct input features, along with their preprocessing, for the specific learning algorithm one wants to use. Eventually, we consider the S&P500 historical time series, predicting trend on the basis of data from the past days, and proposing a novel approach based on combination of wavelets and CNN, which outperforms the basic neural networks ones. We show, that neural networks are able to predict financial time series movements even trained only on plain time series data and propose more ways to improve results.