Machine Learning tools for investing. Does Machine Learning add value in making investment decisions?

Author: Denis Salkovic
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT,

In modern finance machine learning is getting more and more attention due to the massive data increase in recent years. The goal of this thesis was to find out if machine learning is helping investment analysts with making investment decisions. For that goal different components were determined which are essential for investment analysis. These were Performance, Risk, Growth and Cost of Capital. To determine if machine learning is helpful in making investment decisions, it was evaluated how machine learning is useful for these components, which methods it has and how they are performing. Several empirical research articles were evaluated and significant findings of these were filtered to answer the research question. Analyzing these components this study found that machine learning is undoubtedly a helpful assistant for investment analysts and should be considered when dealing with important investment decisions. In some cases, machine learning alone is not as effective as in cooperation with the investment analysts.

Graduation Committee members:

Dr. Robert Gutsche

Dr. Ekaterina Svetlova

Keywords

Investment analysis, Machine Learning, Performance, Risk, Growth, Cost of Capital

1. INTRODUCTION

In today's world more and more data are captured by humans. Every day humans generate and capture more than 2.5 quintillion bytes of data(Ndikum, 2020). This amount especially was generated in the last few years. The number of data gathered every year is expected to exceed 149 zettabytes in 2024 (Arne Holst via Statista, 2021). The prediction and forecasting of asset prices and returns remains one of the most challenging and exciting problems for quantitative finance and practitioners alike (Ndikum, 2020). Predicting returns in the stock market is usually posed as a forecasting problem where prices are predicted. Intrinsic volatility in the stock market across the globe makes the task of prediction challenging (Basak et al., 2019). Due to the massive increase of data this becomes more and more complex (Cao et al., 2021). Therefore, it became increasingly clear that machine learning is a powerful tool for asset management. Machine Learning increased in popularity in recent years thanks to a large amount of qualitative financial data sets have become available (Wu et al., 2019). Driven by this increase in computing power and data researchers and investment firms point their attention to techniques like Machine Learning (Ndikum, 2020). This technique (Machine Learning) is much more frequently used where the amount of data is so big that humans lose their overview. Machine learning algorithms can potentially overcome several limitations with the extant models and process financial statement information more efficiently in forecasting future earnings since they can efficiently handle high dimensional data (Cao et al., 2021). Also, in finance Machine Learning is seeing increased adoption in predictive techniques (Bracke et al., 2019). Especially in order to make profit of the security market, prediction models of the stock market are of important nature. It is clear that the price movement in the stock market is well studied by investors in the hopes of making profit and getting above average returns. It is also clear that the stock market is dynamic in nature (Pal Taneja, 1977). Often it is known which factors influence the stock market and the share prices but usually it is not possible to predict it (Pal Taneja, 1977). Investors could increase their return on investment if it becomes possible to predict future stock prices (Nikou et al., 2019). In the past often traditional methods were used as forecasting methods such like the CAPM or Fama French Three Factors model. Recently, also methods of Machine Learning found their way into official business statistics. These were experimentally tested in the case of classification decisions and the generation of new data (Dumpert & Beck, 2017). Having introduced the relevance of machine learning the aim of this research is to find out if machine learning does help the investment analysts with investment analysis and leads me to the research question: Does machine learning add value in making investment decisions?

1.1 Research objective and question

This research aims to identify if Machine Learning helps investment analysts with investment decisions. Hence, the performance of machine learning is evaluated. The primary objective is to assess where Machine Learning is helping to make accurate forecasts for stock market prices and investments. At first, I will give a brief introduction into investment analysis and machine learning. After having these introduced, I will thoroughly dig deeper into relevant literature to find where machine learning comes into play to help investment analysts making forecasts more accurate. Therefore, the hypotheses of this research are:
1. Machine learning is associated with predicting performance of investments.
2. Machine learning is associated with predicting growth.
3. Machine learning is associated with predicting risk of investments.

4. Machine learning is associated with predicting cost of capital.

To test these hypotheses useful components for analyzing performance of machine learning will be later determined. After analyzing these components, the findings should help with the evaluation of machine learning performance and if investment analysts should consider using machine learning techniques as an assistant for their work.

2. LITERATURE REVIEW

2.1 Investment Analysis

To analyze if machine learning is a helpful assistant for making investment decisions, I will briefly explain what investment analysis all is about. Investment analysis is defined as the process of evaluating an investment for profitability and risk. In investment analysis an analyst is measuring how a given investment is a good fit for a portfolio. It involves researching and evaluating a security, for example a stock price, to predict its future performance and determine if it's worth to add a specific investment to a portfolio. Investment analysis covers more aspects than its name implies. Investment analysis is a broad term that encompasses many different aspects of investment. It can come in handy for predicting about future returns. To make predictions about future returns business executives can examine past returns. The predictions open up to the type of investment vehicle which remains in best interest of the needs of an investor or evaluating securities such as stocks and bonds for valuation and investor specificity. Investment analysis can facilitate how an investment is likely to be executed and how great the opportunity is for a given investor. Investment analysis is essential to any sound portfolio management strategy (Magalie, n.d.). As investors are often not familiar with investment analysis, they´re often hiring investment analysts. Making investment decisions requires thorough analysis and investigation into investment. Investors can use multiple analysis approach to find out the most effective method. The approach could be a bottom-up investment analysis approach or top-down approach. Bottom-up investing utilizes a microeconomic approach to investing most importantly rather than a macroeconomic one. Top-down investment analysis focuses on economic, market and industrial trends before making a more granular investment decision (Magalie, n.d.). Put simply how Warren Buffet once said, an investment is nothing less than selling at a higher price than bought. Hence, the investment decision is all about analyzing price and market value. If price is lower than market value stocks are bought and sold if price exceeds market value.

2.2 Machine learning in finance

Basically, Machine learning is artificial intelligence that uses statistical models to make predictions. In finance, machine learning algorithms are mainly used for the purpose of detecting fraud, automating trading activities and providing advice and recommendations to investors. Machine learning algorithms can efficiently handle high dimensional data. The generation of corporate earnings is a complex process involving numerous business transactions. In contrast to traditional linear models, machine learning algorithms can accommodate more complex and subtle relationship between financial statement line items and future earnings. Economic theories and empirical evidence suggest the existence of nonlinear relationships between financial statement line items and future earnings (Cao et al., 2021). In theory a deep neural network can find the relationship for a return, no matter how complex and non-linear. This is a far cry from both the simplistic linear factor models of traditional financial economics and from the relatively crude, ad hoc methods of statistical arbitrage and other quantitative asset

management techniques (Heaton et al., 2016). Recently, many researchers have demonstrated the impressive empirical performance of Machine Learning algorithms for asset price forecasting when compared with models developed in traditional statistics and finance (Ndikum, 2020). Machine learning algorithms learn from historical data in a process known as training and subsequently make accurate predictions on new data. Samuel (1959) defined machine learning as a chance to not implement problem solving methods as exact as before: The programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning. Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort (Samuel, 1959). Gu and colleagues use the term of machine learning to describe a diverse collection of high-dimensional models for statistical prediction, combined with so-called "regularization" methods for model selection and mitigation of overfit and efficient algorithms for searching among a vast number of potential model specifications (Gu et al., 2020). The high dimensional nature of machine learning methods enhances their flexibility relative to more traditional econometric prediction techniques. This flexibility brings hope of better approximating the unknown and likely complex data generating process underlying equity risk premiums (Gu et al., 2020).Although machine learning algorithms are designed to handle high dimensional data, the inclusion of many irrelevant features increases the risk of overfitting. Thus, a set of sufficiently disaggregated financial statement line items without overwhelming the algorithms with excessive irrelevant noise needs to be selected (Cao et al., 2021).

2.2.1 Supervised machine learning

Supervised machine learning is the construction of algorithms that are able to produce general patterns and hypotheses by using externally supplied instances to predict the fate of future instances. Supervised machine learning classification algorithms aim at categorizing data from prior information (Hoda et al., 2016). Whilst there are many types and classes of Machine Learning algorithms a high percentage of the research papers in the current academic literature frame the problem of financial asset price forecasting as a supervised learning problem. In general, the data is decomposed into a training set and a test set. In the training phase the algorithms will learn to approximate the function to produce a prediction. In Machine Learning the performance of algorithms are evaluated using an accuracy measure which is usually a function of the error terms in the test set. Supervised learning is split in two standard formulations: classification and regression. While in classification the learner is required to classify the probability of an event or multiple events occurring in regression the learner is required to predict a real number as the output. In classification the probability of an economic recession or a boom at the stock market can be predicted. In regression annual returns of a financial asset in the future or a stock price is predicted (Ndikum, 2020). Supervised machine learning is the construction of algorithms that can produce general patterns and hypotheses by using externally supplied instances to predict the fate of future instances. Supervised machine learning classification algorithms aim at categorizing data from prior information (Hoda et al., 2016). In supervised learning human interaction is needed since the machine learning algorithms are using labeled data, which need to be labeled by humans. Often used supervised machine learning tools are Random Forest and gradient boosting trees.

2.2.2 Unsupervised machine learning

Unsupervised machine learning uses algorithms to analyze and cluster unlabeled data. These algorithms discover hidden patterns or data groupings without the need for human intervention. Unsupervised machine learning algorithms are utilized for three main tasks: clustering, association, and dimensionality reduction. Often used unsupervised machine learning tools are hierarchical clustering, k-means, mixture models, DBSCAN, and OPTICS algorithm.

2.2.3 Deep learning

Deep learning is a type of machine learning methods based on learning data representations. It is an advanced technique of machine learning based on artificial neural network algorithms. As a promising branch of artificial intelligence, deep learning has attracted great attention in recent years (Huang et al., 2020). According to Wu et al. there are two major approaches to forecast equity returns with deep learning: Artificial Neural Networks (ANN) as a blackbox for standard "factors" and forecasting the price time-series. Approach 1 represents the line of thought that feature engineering is critical in building machine learning models, whereas approach 2 represents the mindset that deep learning can automatically extract features so effort on feature engineering should be avoided (Wu et al., 2019).
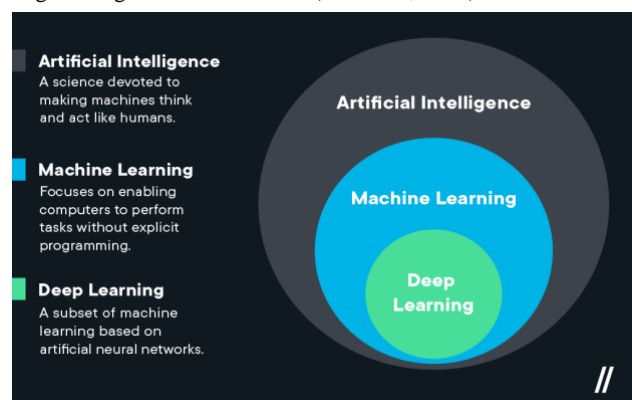

Figure 1: Differences between AI, ML and DL

2.3 The stock market

The most glamorous of all the financial markets is the stock market, which never fails to conger up images of high rollers frantically buying and selling stock and making millions in returns. Unfortunately, the reality of the natures of the market paints a less optimistic picture (Y.-F. Wang, 2003). The financial forecasting or stock market prediction is one of the hottest fields of research lately due to its commercial applications owing to the high stakes and the kinds of attractive benefits that it has to offer (Majhi et al., 2007). Unfortunately, stock market is essentially dynamic, non-linear, complicated, nonparametric, and chaotic in nature (Tan et al., 2005). This makes it extremely hard to model with any reasonable accuracy (Y.-F. Wang, 2003). The time series are multi-stationary, noisy, random, and has frequent structural breaks (Oh & Kim, 2002; Y.-F. Wang, 2003). Furthermore, movements of the stock market are affected by several economic factors like political events, firms´ policies, bank rate, bank exchange rate, general economic conditions, movements of other stock market and so forth (Yudong & Lenan, 2009). Although people have come up with many methods do try and do so, traditionally the best performers have been speculators who use their considerable knowledge of the market to predict the next trend. As these speculators are only human, they are limited in their capacity to assimilate information and spot subtle trends in the information, which may be the indicators of an impending change in the value of the stock market (Y.-F. Wang, 2003). On the contrary site, Usmani et al. propose that machine learning techniques are capable of predicting the stock market performance (Usmani et al., 2016).

3. FINANCIAL FRAMEWORK

To analyze if machine learning is a helpful assistant for investment analysts, I will choose several investment analysis components that are useful to determine if machine learning is helping with investment decisions. These components are Performance, Growth, Risk, and Cost of Capital. With the help of previous research, I will build the theoretical framework around these components. The framework is about finding out if machine learning is associated with the four components and if ML then helps with investment decisions. Why these components? Performance, Growth, and cost of capital are part of the Value formula for investments which is performance divided by cost of capital minus growth. Risk is always of importance as almost every investment contains some degree of risk.

3.1 Performance

The performance component is quite broad since all the other components at the end will result in a performance forecast of investments. Therefore, the performance component is the most meaningful to make an investment decision. Performance will be evaluated according to different performance measures. On the one hand it will be looked at the Mean Squared Error of fundamental analysis compared to machine learning techniques. Furthermore, the F-score is used in some of the literature as a performance criterion. Also, different machine learning techniques are compared to each other to evaluate which algorithm is performing best. Several authors analyzed different kinds of markets. Hence, in the findings section, not only the S&P 500 will find but also the Indian stock market, Taiwan stock market and other markets. It will be evaluated if machine learning techniques result in a significant performance improvement over the classical methods of the investment analysts. As there are several machine learning algorithms in the findings sections the performance of some of them will be compared to each other and evaluated.

3.2 Growth

To forecast share prices of several firms the predictions of future sales are essential as these are very important to determine the growth potential of a firm. Since the share price of a firm is dependent on the performance of for example their sales it is of importance to predict future sales. If a firm has a promising sales forecast the demand of this firm's shares is probably increasing and as demand increases supply decreases which consequently results in a rising share price. Therefore, sales prediction is an essential part of the growth component. Most of business organizations heavily depend on a knowledge base and demand prediction of sales trends. The accuracy in sales forecast provides a big impact in business. Data mining techniques are very effective tools in extracting hidden knowledge from an enormous dataset to enhance accuracy and efficiency of forecasting (Cheriyan et al., 2018). On the counter side of growth stands the decline which is also a very important indicator for investments. Even at the beginning of 2008, the economic recession of 2008/09 was not being predicted by the economic forecasting community. The failure to predict recessions is a persistent theme in economic forecasting (Nyman & Ormerod, 2016). Hence, it is aimed to look in the literature if findings are available of machine learning helping with the prediction of future sales growth and predicting recessions to avoid pitfalls in investing.

3.3 Risk

The unanticipated part of the return, that portion resulting from surprises, is the true risk of any investment. The risk of owning an asset comes from unanticipated events, otherwise the investment would be perfectly predictable (Hillier et al., 2017).

Risk can basically be differentiated in two types of risk. First, Systematic risk or market risk, which is a risk that influences a large number of assets. Systematic risk is often measured with the beta coefficient, which is defined as the amount of systematic risk present in a particular risky asset relative to that in an average risky asset. Second, Unsystematic risk or also called asset-specific risk, which is a risk that affects at most a small number of assets. This risk is, as it often only affects a small number of assets, diversifiable when building a portfolio with more assets than only one. In finance, systemic risk is a crisis that leads to the collapse of an entire financial system or entire market of an area or country, even global markets. The greatest impact of the global financial crisis in 2008, with strong economic destructive power and causing a huge chain reaction to destroy the financial industry, enabled systemic risk to be regarded as a crucial factor in relation to financial safety. Hence, over the past decade, a large amount of ground-breaking academic research has focused on systemic financial risks, including the study of the financial ecosystem, financial supervision, etc. Because of risk being hidden in modern large-scale financial systems, intelligent and automatic machine learning methods become a concerned tool to assess and detect the systemic risk from increasingly complex financial network, big data of financial transactions and market sentiments (Kou et al., 2019).

3.4 Cost of Capital

The cost of capital is defined as the minimum required return on a new investment. It is called this because the required return is what the firm must earn on its capital investment in a project just to break even (Hillier et al., 2017). Estimating a firm´s implied cost of capital is a key issue in accounting and finance. Prior researchers have largely relied on time-series estimates and/or analyst forecasts as inputs into the implied cost of capital calculation. The expected return is a major topic in research related to firm resources allocations or decision making. Since the ex-ante expected returns are unknown, in the past, researchers use historical realized returns to forecast expected returns based on time-series models, but academic researchers indicate that the expected returns by time-series models do not work well, majorly because ex post realized returns are noisy. There are some important limitations to these approaches, including limited data availability on analysts' forecasts and substantial noise when relying on time-series estimates. Deep learning techniques had been established in recent time on common accounting information to offer substantial improvement to the earnings prediction model (X. Wang, 2020).

4. METHODOLOGY

This research aims to answer the following research question: Does machine learning add value in making investment decisions? To be able to answer this question, several relevant empirical literatures will be studied on if machine learning is helpful in predicting the performance, growth, risk, and cost of capital in the upcoming future to make appropriate and profitable investments. This research will be conducted in the form of a literature review and a meta-analysis. Meta analyses are used to combine statistical analyses of multiple scientific studies. In other words, Meta-Analysis is the statistical procedure for combining data from multiple studies. And as multiple studies are addressing the question if Machine Learning is helpful in finance the conditions for a meta-analysis and a literature review were optimal. Significant papers will be put into a table where the findings of different relevant papers concerning the determined components will be explained in more detail in the findings section. Papers were conducted from the library of Utwente, the Journal of Finance and Google scholar. The most popular articles were used, meaning the first papers appearing on the list of the search query. Eight Papers were used for the

Paper	Performance	Growth	Risk	Cost of Capital
Ndikum	Ndikum finds that Performance of machine learning outperforms fundamental analysis. Mean squared error of ML techniques are significantly lower than of the CAPM. The highest MSE of a ML technique was 0.3628 while CAPM had 1.6001.	x	x	x
Yudong et al.	Yudong et al. find that their IBCO-BP (improved bacterial chemotaxis optimization) model performs significantly better than the often-used BP (Back propagation) neural network model. Furthermore, it takes less time for training. Their model had a computation time of 16.2503 per minute and the BP 25.2044 per minute for prediction 15 days ahead.	x	x	x
Milosevic	Milosevic compared several machine learning algorithms on Precision, Recall and F-score and finds that out of the 8 tested algorithms, Random Forest performs the best with 0.751 on Precision, Recall and F-score.	Growth not necessary. Milosevic proposes that past growth is not being necessary for future growth prediction.	x	x
Das et al.	Das et al. compared the two machine learning algorithms BPN with SVM and found that SVM is better than BPN.	x	x	x
Wang	Wang tested the fuzzy rough set system since he thinks neural networks are unsuccessful. After running 180 trials his system was 93% accurate.	x	x	x
Cheriyan et al.	x	Cheriyan et al. tested three different machine learning algorithms for sales prediction. They found that gradient boost algorithm was the most accurate with 98% accuracy.	x	x
Saito et al.	x	Saito et al. looked at sales growth prediction system in the Japanese economy. They found that to some extent their results imply that machine learning techniques are feasible for sales growth prediction. The machine learning algorithms show classification accuracy in the range of 70% ~ 90%.	x	x
Nyman and Ormerod	x	Nyman and Ormerod found that machine learning technique of random forest has the potential to give early warning of recessions.	x	x
Gu et al.	Gu et al. found that machine learning methods are the most valuable for forecasting larger and more liquid stock returns and portfolios.	x	Gu et al. found that machine learning has great potential for improving risk premium measurement, which is fundamentally a problem of prediction	x
Nabipour et al.	Nabipour et al. found that deep learning methods with binary data performed the best in predicting stock market movement. With continuous data the algorithms had an average F-score of 63%, with binary data up to 83%.	x	Nabipour et al. found that machine learning and deep learning significantly can reduce the risk of trend prediction.	x
Wang X.	x	x	x	Wang found that a deep neural network can be used productively in the context of earnings forecasting and uncovering a firm´s implied cost of capital
Usmani et al.	The Multi-Layer Perceptron algorithm of machine learning predicted 77% correct market performance.	x	x	x

Table 1: Significant Findings

Performance component, four Papers for the Growth component, two papers for the Risk component and one for the cost of capital component. Therefore, the nature of this research is quantitative.

After this research it should become clearer if machine learning is a helpful assistant in terms of making investment decisions regarding the given components. In the significant findings part the findings of the different authors are briefly summarized in Table 1. In the following findings section the before determined components are elaborated in more detail with suggested methods they used.

5. SIGNIFICANT FINDINGS
See table 1.

5.1 Summary of Table
For the performance component all the authors found that machine learning is very precise in predicting performance of the stock market. The accuracy of the predictions ranges from 63% to 93%. Additionally, the MSE of machine learning algorithms is significantly lower than of investment analysts. The difference between the worst ML algorithm´s MSE and the best analyst´s MSE is 1,2373.

For the Growth component all articles used show that machine learning is better able to predict future growth. The accuracy ranges from 70% to 98%. Furthermore, Machine learning seems to be a promising technique for giving early warning of economic recessions.

The two articles considered for risk were both stating that machine learning can significantly lower the risk of trend prediction and improving risk premium measurement. Unfortunately, no relevant numerical values like beta were found for machine learning.

For cost of capital the only author considered found that deep learning techniques can lead to earnings forecasts which offer marginal information content beyond the ability of human financial analysts. Numbers for this are considered in 6.4.

6. FINDINGS
6.1 Performance of ML algorithms
As Performance is the most important and most meaningful component it contains the most findings in the table. As the table only briefly summarizes the findings of the different authors, in this section their findings will be elaborated on in more detail. Ndikum (2020) researched in his paper the performance of machine learning algorithms and techniques that can be used for financial asset price forecasting. He compared different supervised machine learning algorithms with the fundamental analysis tool CAPM. For the performance metric he used the Mean Squared Error (MSE). The results demonstrate that the six machine learning algorithms he used (NGBoost, XGBoost, Catboost, LightGBM, Shallow FNN, Deep FNN) significantly outperform the CAPM model. The Catboost model had the lowest MSE of the machine learning models with 0.3125 and the Shallow FNN the highest with 0.3628. Whether it is the gradient boosting trees or the neural networks, all significantly outperformed the CAPM which had an MSE of 1.6001. This results in superior performance of the supervised machine learning techniques to forecast annual returns.

Yudong & Lenan (2008) compare their improved bacterial chemotaxis optimization (IBCO) machine learning algorithm with the often used back propagation neural network (BPNN) for stock index prediction.
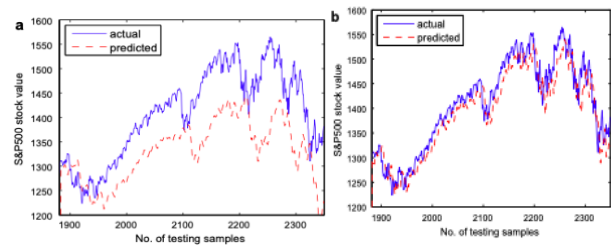

Figure 2: Comparison of Yudong´s & Lenan´s model (b) and traditional BP model (a) for prediction fifteen days ahead

As it can be seen on the graphs, the IBCO developed by Yudong & Lenan is significantly more accurate than the traditional BP model in predicting the S&P500 index fifteen days ahead. Furthermore, they tested that the computation time of the IBCO model is significantly outperforming the BP model. These findings suppose that their IBCO model offers less computational complexity, better prediction accuracy and less training time.

Comparison of computation time between IBCO model and BP.

Days ahead	1	3	5	7	9	15
Time of IBCO/min	15.4983	15.6306	15.7734	15.9040	16.1438	16.2503
Time of BP/min	24.3279	24.4821	24.6203	24.7955	24.9186	25.2044

Table 2: Yudong´s & Lenan´s Comparison of computation time between IBCO model and BP

Milosevic presents in his paper a machine learning aided approach to evaluate the equity´s future price over the long term. He selected 28 financial indicators and then tested eight different machine learning algorithms with Precision, Recall and F-score and found that Random Forest is performing best. To further improve the results of the random forest algorithm he found that the 28 selected financial indicators need to be reduced to eleven indicators. Hence, the random forest algorithm is even more efficient and faster with smaller set of features (Milosevic, n.d.)

Algorithm	Precision	Recall	F-score
C4.5 decision trees	0.687	0.687	0.687
SVM with SMO	0.639	0.630	0.624
JRip	0.635	0.635	0.635
Random Tree	0.668	0.668	0.668
Random Forest	**0.751**	**0.751**	**0.751**
Logistic regression	0.643	0.637	0.633
Naïve Bayes	0.545	0.530	0.487
Bayesian Networks	0.625	0.618	0.612

Table 3: The results of machine learning based on 28 financial indicators

Das & Padhy (2012) tested two machine learning techniques to predict future prices traded in the Indian stock market. They compared the back propagation technique (BP) and support vector machines (SVM) with each other. They used the normalized mean squared error (NMSE), the mean absolute error (MSE) and directional symmetry as performance criteria.

Futures Details	NMSE		MAE		DS	
	SVM	BPN	SVM	BPN	SVM	BPN
BANK NIFTY	0.929	1.335	0.238	0.250	91.251	83.880
CNX 100	1.012	1.333	0.382	0.381	87.17	85.672
CNX INFRA	1.032	1.471	0.293	0.299	88.192	80.895
S&P CNX 500	1.152	1.378	0.332	0.336	85.671	80.597
S&P CNX NIFTY	1.019	1.100	0.389	0.403	85.17	88.059

Table 4: Comparison of the results of SVM & BPN

SVM provides a smaller NSME and MAE and larger DS than of BPN in most of the cases. The results of their research

demonstrate that forecast for Indian stock market prices of the SVMs is better than of the BPN.

To overcome the in his opinion unsuccessful neural network technique for predicting stronger rules of stock prices Wang (2003) developed a new effective method called fuzzy rough set system. His system has been written with Visual BASIC on an IBM PC and includes two major modules: visual display agent and mining agent. The current price of the stock can be checked through the visual display agent whereas the mining agent provides the user with buy/sell the stock according to the mining results. They retrieved the data from the Taiwan stock market. After running over 180 trials they achieved 93% accuracy with their new developed fuzzy rough set system in predicting the ranking of stock prices.

Gu et al. found that machine learning methods are most valuable for forecasting larger and more liquid stock returns and portfolios. They performed a comparative analysis of methods in the machine learning repertoire. Their findings demonstrate that machine learning methods, at the highest level, can help improve analysts´ empirical understanding of asset prices. Neural networks and regression trees are the best performing methods. Furthermore, the results of their research show that all methods agree on a small set of dominant predictive signals, the most powerful predictors being associated with price trends including return reversal and momentum. The next most powerful predictors are measures of stock liquidity, stock volatility and valuation ratios. Overall, Gu et al. found that machine learning algorithms for return prediction brings promise for both economic modeling and for practical aspects of portfolio choice.

Nabipour et al. researched the performance of the prediction task of stock market movement by machine learning and deep learning algorithms. They employed nine different machine learning algorithms to detect which one is performing best. They supposed two approaches for input values to models, continuous data, and binary data. Their results showed that there was a significant improvement in the performance of models when they use binary data instead of continuous one. The deep learning algorithms (RNN and LSTM) were superior in both approaches. As previous authors they are also using the F-Score as a performance criterion. As you can see in the graphs, the deep learning algorithms are either in continuous data the best performing or in binary data (Nabipour et al., 2020).
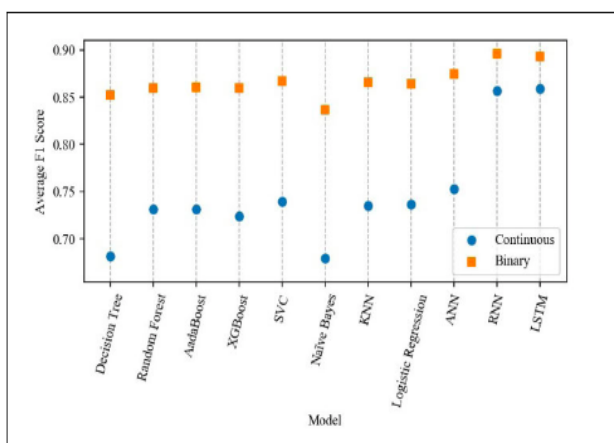

Figure 3: The average of F1-Score with continuous and binary data

Usami et al. (2016) compared performance of different machine learning techniques to predict the Karachi Stock Exchange (KSE). The results of the four algorithms used: Single Layer Perceptron (LSP), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF) and Support Vector Machine (SVM) can be seen

in table 5. SVM performed best on training set while MPL algorithm did well on test data set. Therefore, MLP seems to be more efficient in predicting the market performance. Hence, Usami et al. found that machine learning techniques are capable of predicting the stock market performance. The best performance exhibits the MLP algorithm with 77% correct predicted market performance.

Data set used for verification	Machine Learning Algorithms			
	SLP	MLP	RBF	SVM
Training Set	83%	67%	61%	100%
Test Set	60%	77%	63%	60%
Average	71.5%	72%	62%	80%

Table 5: Comparison of Machine Learning Techniques

6.2 Growth prediction

Milosevic (n.d.) just briefly mentioned something about growth in his paper which is about equity forecast, more precisely about predicting long term stock price movement using machine learning. He found that machine learning does not need information about growth. From this it can be concluded that ratios and information that describes current financial state of the company, without a look at the past performances is enough for predicting future behavior of the company. This is especially useful for investors who want to invest in startup companies. Though Milosevic also admitted that it hast to be further tested in the coming research if companies can be valued and their future can be predicted only by looking at present data.

Cheriyan et al. (2018) researched a different direction of growth. Since most of business organizations heavily depend on a knowledge base and demand prediction of sales trends, which is increasingly difficult for analysts due to the massive increase of data, they investigated intelligent sales prediction using machine learning techniques. They tested three different machine learning techniques. As also in previous mentioned research they are comparing Gradient Boosted Trees (GBT), Decision Trees (DT) and the Generalized Linear Model (GLM).

Model Name	Performance Summary of ML Algorithms				
	Accuracy Rate (%)	Error Rate	Precision	Recall	Kappa[1]
GLM	64	36	5.36	0	0
DT	71	29	11.24	15.61	0.501
GBT	98	2	50	50	0.962

Table 6: Performance summary of sales prediction

After testing the accuracy rate, error rate, precision, recall and kappa Cheriyan et al. concluded that the Gradient Boosted Tree is performing best in all values out of the three machine learning models as it can be seen in table 5. Furthermore, they found that an intelligent sales prediction system is required for business organizations to handle enormous volume of data.

Saito et al. carried out an empirical evaluation of machine learning performance in corporate sales growth prediction. They also tested five different machine learning techniques, which are random forest (RF), weighted random forest (WRF), gradient boosting decision tree (GBDT), support vector machine (SVM) and the least- squares probabilistic classifier (LSPC). As performance metrics Saito et al. chose the accuracy ratio (AR), weighted F1-scores, and the area under the ROC curve (AUC).

They tested four different industries and applied the machine learning techniques to Japanese sample firms. The machine learning algorithms show classification accuracy in the range of 70% ~ 90%.

Software:

	AR	F-score	AUC
RF	0.7345	0.4333	0.7054
WRF	0.7311	0.4326	0.7047
GBM	0.7533	0.3498	0.5969
SVM	0.7761	0.0685	0.6240
LSPC	0.6652	0.3195	0.5684

Service:

	AR	F-score	AUC
RF	0.7597	0.4155	0.7760
WRF	0.7606	0.4149	0.7760
GBM	0.8066	0.3619	0.6356
SVM	0.8447	0.1159	0.6886
LSPC	0.7513	0.3148	0.6188

Table 7: Classification results of 2 out of the 4 industries

Other than in other empirical research presented no single classifier dominated among those that showed the best results in each performance metric. SVM showed the best performance in AR, although RF showed the best performance in F-score and AUC. To conclude, Saito et al. found that to some extent, the results imply the feasibility of a machine learning system for sales growth prediction. Additionally, they remarked that realization of a sales-growth prediction system would encourage investment in Japanese firms and therefore lead to growth of Japanese economy (Saito et al., 2021).

Nyman & Ormerod (2016) analyzed if machine learning algorithms can predict economic recession. This is not specifically a finding which belongs to the growth component but is essentially important and should be considered when talking about growth since it is the opposite of growth at least for some undetermined time. They used a small set of explanatory variables from financial markets which would have been available to a forecaster at the time of making the forecast.
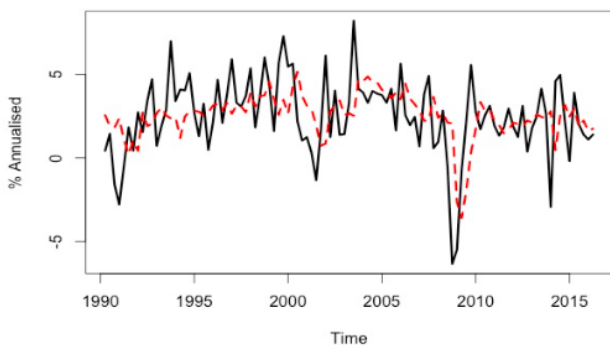

Figure 4: Actual annualized quarter on quarter third estimates US GDP growth and random forest predictions made six quarters previously, 1990Q2-2016Q2

Note: solid black line is actual and the dotted red line the random forest prediction.

As it can be seen on the graph, the random forest predictions would not have got the exact timing of the recession in 2008/09 correct, but a serious recession would have been predicted for early 2009 eighteen month previously. Ormerod and Mounfield (2000) added that machine learning techniques do seem to have considerable promise in extending useful forecasting horizons and providing better information over such horizons. Further research needs to be conducted towards achieving a higher degree of accuracy of prediction.

6.3 Risk prediction of ML algorithms

When selecting investments for a portfolio the risk gets lower the more investments the portfolio contains. But selecting the right investments for a portfolio is not easy. Gut et al. researched empirical asset pricing and performed a comparative analysis of machine learning methods for measuring asset risk premiums. They demonstrated large economic gains to investors using machine learning forecasts. They provide benchmarks for the predictive accuracy of machine learning methods in measuring risk premiums of the aggregate market and individual stocks. The accuracy is measured in two ways: R^2 and the sharpe ratio. After conducting their research Gu et al. found that all the most powerful predictors being associated with price trends including return reversal and momentum. The next most powerful predictors are measures of stock liquidity, stock volatility and valuation ratios. Additionally, they found that with better measurement through machine learning, risk premiums are less shrouded in approximation and estimation error. Therefore, machine learning brings promise for both economic modeling and for practical aspects of portfolio choice since it reduces risk.

Nabipour et al. did not researched something specifically linked to risk, but what he researched leads to risk prevention and therefore I decided to include his findings into the risk component. They conducted a research towards stock market trend prediction via machine learning and found that machine learning and deep learning perform very well when binary data is used. Figure 3 shows this very well. What Nabipour et al. concluded from their research is that machine learning significantly reduces the risk of trend prediction of the stock market.

6.4 Cost of Capital prediction of ML algorithms

Due to lack of the literature not much articles could be found on this component. Wang (2020) applied a deep learning approach to show that the deep learning approach offers incremental explanatory power in predicting future earnings and in estimating the associated implied cost of capital. She saw a lack of research in this specific topic and a problem in the prior often used time-series estimates or analyst forecasts as inputs into the implied cost of capital or expected return calculation. She saw the problem in some important limitations to these approaches since the data availability is very limited on analyst forecasts and substantial noise when relying on time-series estimates. In her results Wang found that deep learning techniques can lead to earnings forecasts that offer marginal information content above and beyond human analyst forecasts. The machines can extract information from observable data that human analysts either miss or process incorrectly. Furthermore, Wang noticed that combining analyst predictions and deep-learning techniques may lead to substantially superior forecasts that either approach used in isolation.

	E_{t+1}			E_{t+2}		
	Deep Learning	Analysts	Difference	Deep Learning	Analysts	Difference
Panel A: Bias of Earnings Forecasts						
Mean	-0.0372	-0.0151	-0.0220	-0.0163	-0.0404	0.0241
	-5.86	-0.49	-0.72	-2.00	-1.39	0.84
Median	0.0035	-0.0066	0.0101	0.0033	-0.0267	0.0300
	2.72	-4.86	5.91	1.05	-6.90	8.23
Panel B: Accuracy of Earnings Forecasts						
Mean	0.1013	0.1016	-0.0003	0.1093	0.1347	-0.0254
	14.76	3.14	-0.01	23.76	4.12	-0.75
Median	0.0303	0.0153	0.0150	0.0449	0.0381	0.0068
	29.57	9.21	7.15	33.55	10.35	1.65
Panel C: Annual Earnings Response Coefficient (ERC)						
Mean	0.0688	0.0488	0.0200	0.2123	0.1332	0.0791
	7.45	4.31	1.33	11.56	5.98	3.30

Table 8: The Bias, Accuracy and Annual Earnings Response Coefficient of Analysts´ Forecasts and Deep Learning´s Forecasts on Earnings

As can be seen in the table deep learning can generate earnings estimates which are less biased and better fit the market expectation, comparing to the linear model as well as financial analysts.

7. LIMITATIONS

Before concluding this research, some limitations need to be mentioned. First, the components were chosen to the existing knowledge and therefore there are a lot of other components which could be considered for the determination whether machine learning is helping analysts with investment decisions. Furthermore, cost of capital component is very limited since not much empirical research has been found on this component. Similar for the risk component. As for the other two components at least four articles were considered, for risk it was only two and for cost of capital only one. To have a better benchmark and a more meaningful result for these components, more empirical articles would be of great help. Additionally, not enough data was found for adding numerical values to the risk component like if machine learning is able to reduce the beta of portfolio investments. This study tried to find an answer on whether machine learning can help investment analysts with investment decisions and additionally gave some examples and comparisons of different machine learning algorithms. To further evaluate where machine learning specifically is helping with certain analyses, more empirical research needs to be done. Additionally, more components could be determined and more specifically applied to machine learning algorithms. Another limitation is that this research is not giving insights which algorithm is performing best for which component.

8. FURTHER RESEARCH

Since this study researched if machine learning is a helpful assistant for investment analysts and answered it with a huge yes, future studies could concentrate on which specific machine learning algorithms could be best used for the single components of investment analysis. This could serve as a guide for investment analysts who are not yet familiar working with machine learning. As this research showed, machine learning is undoubtedly helping investment analysts with investment decisions. But the question which algorithm should be applied for most effective performance of the several components is a subject for future research.

9. CONCLUSION

With the massive increase of data in the upcoming years, it gets more and more complex to analyze the financial data efficiently. Investment analysts alone can´t handle all the data by their own to come to good investment decisions. Therefore, the time of machine learning is still emerging.

The purpose of this research was to give organizations and individuals an insight on investment analysis and how machine learning is positively influencing it. For these who were considering applying machine learning techniques this research should serve to help make the decision easier. Most of the authors of the research paper used in this study aren't questioning whether machine learning is a helpful assistant for investment analysts to make investment decisions. They are rather comparing different machine and deep learning algorithms to find out which is performing best. Therefore, the research question is answered: Machine learning is an undoubtedly helpful assistant for investment analysts since it is outperforming investment analysts in almost all components. And the hypotheses came out to be confirmed as Machine learning is associated with all the determined investment analysis components. Out of this research the question rather arises which machine learning algorithm should be applied for which component. Additionally, it was found that machine learning is outperforming investment analysts in accuracy of forecasts, precision and is also attempting less mistakes.

To sum up, machine learning´s performance in all the determined components is superior to the investment analysts. In some approaches the best solution is to combine the machine learning algorithms with the investment analysts to keep the performance of the investments at a maximum.

10. ACKNOWLEDGEMENT

11. REFERENCES

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *North American Journal of Economics and Finance*, *47*, 552–567. https://doi.org/10.1016/j.najef.2018.06.013

Bracke, P., Datta, A., Jung, C., & Sen, S. (2019). *Staff Working Paper No. 816 Machine learning explainability in finance: an application to default risk analysis*. www.bankofengland.co.uk/working-paper/staff-working-papers

Cao, K., You, H., Bhattacharya, U., Cao, S., Chen, K., Chen, P., Chen, Z., Huang, A., Hung, M., Morris, A., Li, K., Li, X., Murray, S., Reeb, D., Sloan, R., Wang, R., Zang, A., & Zheng, Y. (2021). *Fundamental Analysis via Machine Learning*.

Cheriyan et al., Institute of Electrical and Electronics Engineers, & Institute of Electrical and Electronics Engineers. United Kingdom and Republic of Ireland Section. (2018). *Intelligent Sales Prediction Using Machine Learning Techniques*.

Dumpert, F., & Beck, M. (2017). Einsatz von Machine-Learning-Verfahren in amtlichen Unternehmensstatistiken. *AStA Wirtschafts- Und Sozialstatistisches Archiv*, *11*(2), 83–106. https://doi.org/10.1007/s11943-017-0208-6

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Heaton, J. B., Polson, N. G., & Witte, J. H. (2016). *Deep Learning in Finance*. http://arxiv.org/abs/1602.06561

Hillier, D., Clacher, I., Ross, S., Westerfield, R., & Jordan, B. (2017). *Fundamentals of Corporate Finance* (Third Edition). Mc Graw Hill Education.

Hoda, M. N., Bharati Vidyapeeth's Institute of Computer Applications and Management (New Delhi, I., Institute of Electrical and Electronics Engineers. Delhi Section, & International Conference on Computing for Sustainable Global Development (3rd : 2016 : New Delhi, I. (2016). *A review of Supervised Machine Learning Algorithms*.

Huang, J., Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. In *Frontiers of Business Research in China* (Vol. 14, Issue 1). Springer. https://doi.org/10.1186/s11782-020-00082-6

Kou, G., Chao, X., Peng, Y., Alsaadi, F. E., & Herrera-Viedma, E. (2019). Machine learning methods for systemic risk analysis in financial sectors. In *Technological and Economic Development of Economy* (Vol. 25, Issue 5, pp. 716–742). Vilnius Gediminas Technical University. https://doi.org/10.3846/tede.2019.8740

Magalie, D. (n.d.). *Introduction to Investment Analysis and Portfolio Management*. Https://Mgtblog.Com/Introduction-to-Investment-Analysis/.

Manoppo, C. P. (2015). The Influence of ROA…. *Jurnal EMBA*, *691*, 691–697.

Milosevic, N. (n.d.). *Equity forecast: Predicting long term stock price movement using machine learning*.

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis. *IEEE Access*, *8*, 150199–150212. https://doi.org/10.1109/ACCESS.2020.3015966

Ndikum, P. (2020). *Machine Learning Algorithms for Financial Asset Price Forecasting*. http://arxiv.org/abs/2004.01504

Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, *26*(4), 164–174. https://doi.org/10.1002/isaf.1459

Nyman, R., & Ormerod, P. (2016). *Predicting Economic Recessions Using Machine Learning Algorithms*.

Oh, K. J., & Kim, K.-J. (2002). *Analyzing stock market tick data using piecewise nonlinear model*. www.elsevier.com/locate/eswa

Pal Taneja, Y. (1977). Arumugam (1996) showed market anomaliesfor CAPM Fama French. In *Jagadeesh*. Bhandari.

Saito, M., Ohsato, T., & Yamanaka, S. (2021). An empirical evaluation of machine learning performance in corporate sales growth prediction. In *JSIAM Letters* (Vol. 13).

Samuel, A. L. (1959). *Some Studies in Machine Learning Using the Game of Checkers*.

Tan, T. Z., Quek, C., & Ng, G. S. (2005). *Brain-inspired Genetic Complementary Learning for Stock Market Prediction*.

Usmani et al. (2016). *Stock Market Prediction Using Machine Learning Techniques*. IEEE.

Wang, X. (2020). *The Implied Cost of Capital: A Deep Learning Approach*. https://ssrn.com/abstract=3612472

Wang, Y.-F. (2003). *Mining stock price using fuzzy rough set system*. www.elsevier.com/locate/eswa

Wu, Q., Zhang, Z., Pizzoferrato, A., Cucuringu, M., & Liu, Z. (2019). *A Deep Learning Framework for Pricing Financial Instruments*. http://arxiv.org/abs/1909.04497

Yudong, Z., & Lenan, W. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Applications*, *36*(5), 8849–8854. https://doi.org/10.1016/j.eswa.2008.11.028