# Machine Learning and Deep Learning Based Methods for Stock Market Crisis Forecasting

Research Proposal for Bachelor's thesis

by

Nikoloz Jangisherashvili

# Introduction

The stock market has been known as a “barometer of the economy” meaning its fluctuations are closely related to recession or prosperity of the economic market and can be used as an indicator for these events. Stocks nowadays are one of the more easily accessible financial products for financial investors, a solution to capital scarcity for publicly trading companies, and the main business for investment banks and securities companies. Recently people have been showing more and more enthusiasm for applying artificial intelligence to the financial domain through the means of machine learning and deep learning. Along with the development of computer science technology and stronger computational power new possibilities have emerged for applying big data technologies and complex artificial intelligence algorithms for unveiling the mysteries behind the stock market movements. Stock trend, price, and crisis predictions are all the acts of making judgments on value for stocks or conditions in which the stock market will be in for the upcoming future. As the stock market is affected by not only technical variables but also numerous micro and macro factors, such as political or economic indicators or conditions, worldwide news, wars, and global pandemics it is becoming increasingly hard to make any type of predictions on such markets but it also indicates that stock market state can be used as a metric for economic markets.

The Efficient Market Hypothesis is highly accepted in the financial academic domain; the real-world market can give a different perspective and challenge the hypothesis of unpredictability. The importance of stocks for the economic market makes it a meaningful task to tackle and as mentioned above the development of technology has motivated large companies to devote tremendous resources to exploring and exploiting artificial intelligence in this field.

This study will focus on making predictions on stock market crises and more specifically on Dow Jones Industrial Average (DJIA) which is a stock market index that tracks 30 large blue-chip companies publicly trading on New York Stock Exchange (NYSE) and Nasdaq. The DJIA is the second oldest U.S market index and was created by Charles Dow along with his business partner Edward Jones in 1896. The purpose of its creation was to be a proxy for the health of the U.S economy. In the early 20th century, soon after the index was created, the performance of industrial companies was typically closely related to the overall growth rate of the economy and as the first 12 companies in the Dow index were in the industrial sector it cemented the relationship between the index performance and the overall health of the economy. The economy changes over time and so does the DJIA index as the components of the index may be dropped if the company becomes less relevant to current trends of the economy and the replacement will be a better reflection of the shift.

The above-mentioned information serves as an explanation as to why the predictions made on this index are beneficial for making overall assessments of the economy.

## Literature

As mentioned above stock markets have long been an interest of many scholars. The Efficient Market Hypothesis will play a big role in understanding the financial data and the limitations of the predictive power of such models. The hypothesis was introduced in a paper, published in the journal of Business, *The Behavior of Stock-Market Prices* (Fama, 1965). Despite the chaotic nature of stock markets various methodologies have been developed to perfect the forecasting process. In 1952, Markowitz, an American economist raised the concept of portfolio selection in his paper *Portfolio Selection* (Markowitz, 1952). He introduced two investment indices, the variance and the mean of portfolio assets and he formalized investor preferences mathematically. The concept of diversification is very important to understand even when developing a model that will have to work on a single stock index as the mathematical foundations behind diversification are what lays the foundations for the technical analysis of the stock markets for portfolio selection and systematical management. Another important piece of literature for understanding market fluctuations is the 2002 paper *News Sensitive Stock Trend Prediction* (Fung et al., 2002). Using hypothesis testing and a piecewise segmentation algorithm they attempted to predict trends of the time series by categorizing the slopes of each piecewise segment. Related experiments have been conducted on the DJIA, FTSE 100 index and Nikkei 225 index by attempting to forecast stock markets by using wavelet transforms and Recurrent Neural Networks in 2011 (Hsieh et al., 2011), proving that their model worked great on these indices and could be used to earn profits. Further source maps will include the related papers for the development of the algorithms used in this particular research such as *Long Short-Term Memory* (Hochreiter & Schmidhuber, 1997).

## Methodology

The study will revolve around Multivariate Time Series Classification for crisis forecasting, development of different machine learning models, and comparison of these algorithms as well as their results in being able to identify DJIA index price fluctuations before their occurrence. The default target for the model will be identifying the crisis which is currently defined as three consecutive days of adjusted close price percentage drop of at least 5% in the upcoming four trading weeks (20 trading days). These types of falls often result in short-term crises and panics in the

market with a potential to transform into a large-scale crashes. Given that the Time Series data used for making such predictions are complicated with multiple features creating complex connections within each other Neural Network approach should be an effective solution. Artificial Neural Networks (ANN) is one of the most promising techniques for making predictions on complex datasets without prior learning of the relationships between input and output variables which uses feed-forward neural network and back-propagation for weight training based on gradient-descent, creating embeddings in the latent space used for simplifying data representations with a purpose of finding patterns, which is a method for feature-engineering automation. The form of ANN that will be the most promising for this case will be Recurrent Neural Network (RNN). While this algorithm is a robust deep learning technique for dealing with sequential data, usage of vanilla RNN implies that time series follows a pattern that does not require much of long-term connections and thus eliminates the problem of vanishing gradients the RNN faces. On the other hand, Long, Short-term memory (LSTM) and Gated Recurrent Units (GRU) are more suitable for learning from past experiences to make classify, process, and forecast time series with the arbitrary size of time steps. Other than Neural Networks, Support Vector Machine (SVM) has also been widely used for Time Series data gaining trust among scholars, thus it will also be used to classify stock market crises and compare the results with the RNN.

Data used for training the model has so far been separated into two groups: daily trading data and technical indicators. Daily trading data includes Open, close and adjusted close prices as well as the highest price and the lowest price throughout the trading day. Traded volume may also have predictive power and combinations of the above-mentioned data may have a potential usage as well. As for technical indicators, the relative strength index (RSI) may be used as a feature itself or possibly Smoothed relative strength with different versions of the smoothing period. Moving averages are widely used for identifying stock support and resistance levels and may hold predictive power as well. These features will have to be experimented on and tried out in different forms before they become the final input variables.

The research will follow a following working plan: throughout the upcoming weeks the target for the model will be examined, whether it should remain the same or be changed preferably through literature reference or perhaps with a data-driven approach according to historical data not only for DJIA but other indices as well. Once the target is well-defined, the selection process for the input variables will come to pass through a historical analysis of dependencies between the input and output variables. The analysis will also result in the correct choice of training, validating and testing

periods for the models. Concurrently further research and building processes for the Deep and Machine Learning models will take place.

# Bibliography


Cao, L., & Gu, Q. (2002). Dynamic support vector machines for non-stationary time series forecasting. *Intelligent Data Analysis*, *1*, 67–83. https://doi.org/10.3233/ida-2002-6105

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, *1*, 34. https://doi.org/10.1086/294743

Fung, G. P. C., Yu, J. X., & Lam, W. (2002). News Sensitive Stock Trend Prediction. In *Advances in Knowledge Discovery and Data Mining* (pp. 481–493). Springer Berlin Heidelberg. http://dx.doi.org/10.1007/3-540-47887-6_48

Ganti, A. (2003, November 26). *Dow Jones Industrial Average (DJIA) Definition*. Investopedia; Investopedia. https://www.investopedia.com/terms/d/djia.asp

Haykin, S. S. (2009). *Neural Networks and Learning Machines*. Prentice Hall.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *8*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hsieh, T.-J., Hsiao, H.-F., & Yeh, W.-C. (2011). Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing*, *2*, 2510–2525. https://doi.org/10.1016/j.asoc.2010.09.007

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *1*, 77. https://doi.org/10.2307/2975974

Vapnik, V. (2013). *The Nature of Statistical Learning Theory*. Springer Science & Business Media.