Final Project

Big Data Analytics – 55910

Do Macroeconomic Indicators Help Predict Stock Market Trend?

Prof. Nicole Adler

Prof. Lev Muchnik

Dr. Zvi Benami

Ms. Svetlana Shishov

Submit by

Majdal Hindi 300991890

Contents

Executive Summary

This work was done as part of our final project in the field of Big Data Analytics and in collaboration with the risk management department of Clal Insurance. Clal owns insurance companies, agencies, pension funds, provident funds, and other financial instruments. The company is the fourth largest insurance company in Israel (BDI, 2021), the volume of assets managed by the company is estimated at 286 billion NIS (Clal Insurance Ltd, 2021).

We chose this project because its challenging and we believed that it would add to our knowledge in the field of finance and would give us new skills in the field of data analysis. Our research tries to model the S&P 500 price index to predict the future changes.

Clal presented to us few business issues to research regarding to this topic:

- what are the most influential economic indicators affecting S&P 500 stock index performance.
- creating a model based on those factors to predict the future growth of S&P 500 stock index
- realizing if this model could predict stock market crashes.

In the preliminary study, we conducted an in-depth literature review and found several important points that need to be addressed in determining the research methodology. For example, the relationship between macroeconomic indicators and stock prices is confirmed in the most academic works, although there is a lack of comprehensive assessment of causality and dependence of macroeconomic indicators and stock market regarding the time and changing macroeconomic processes (Donatas Pilinkus, 2010). The random walk hypothesis, a theory states that the prices reflected in the stock market are determined by random events independent of the past, i.e., there is no reliable orderly pattern. Therefore, it assumes that a stock's past movement or trend cannot be used to predict its future movement (Godfrey 1964). We also found interesting research argued that if the market is efficient, i.e., the stock price reflects all the changes once they happen, then there is no need for analysts and a "monkey with eyes covered" will be a good investor like an investment man. (Malkiel, 1973). The intention that random selection would be just as good as the selection of professionals in the field, he meant traders in day trading and not those who invest in the long term. We also found that macroeconomic variables provide a statistically significant effect on stock prices in the long run, although the intensity of the effect may vary substantially between different sectors (Prazak's 2018).

Our data was collected from yahoo finance website to obtain the closing price of the S&P 500 index Fred website to obtain **37** US economic indicators that can potentially influence financial market performance by using python programming language, the analysis covered the period between 1986-12-31 / 2021-12-31. After testing correlations and preforming in-depth analysis, we chose the most influential indicators from different economic categorizations to build our model: **Money Supply** - Velocity of M1 Money Stock, **Prices and Inflation** - 5 year / 10-year breakeven inflation rate and consumer price index, **GOV-T** debt and federal debt as precent of gross domestic product, **Income and Expenditure** - Personal Saving Rate, Household Financial Bonds and **Others** - Composite Home Price Index, CBOE Volatility Index, Dividend Ratio.

We have tried several models in order to understand which model succeeds in performing prediction in the best and most accurate way. Eventually, we chose the two leading models: LSTM and Prophet. <u>Prophet</u> is designed by

Facebook; is a time series forecasting library that tries to capture the seasonality in the past data and works well on a large dataset. LSTM is a deep learning model uses neural networks, it's extremely powerful in time series analysis. The prediction results of LSTM model were slightly higher than in Prophet model according to the accuracy metrics, i.e., MAE - 5% error of LSTM compared to 10% of Prophet. The results shows that LSTM model with 50 hidden neurons can provide a superior fit and high prediction accuracy compared to Prophet.
The proposed model can be easily customized to apply in other broad market indices where the data exhibits a similar behavior. Our model can contribute to better understanding of the changes of direction and to expanding it into the world of risk management and scenario analysis. So that Interested investment companies can use the proposed model to better inform the market situation before making their investment decisions. Finally, it should be noted that this is only a basic, starting point model and if Clal or a different risk management companies want to use it, it must be further developed.

Introduction

Clal Insurance Ltd. was founded in 1987 and ranked as the fourth insurance company in Israel (BDI, 2021) with market shares of 18% (total premiums and benefits) and a volume of assets of 286 billion NIS as of 2021. The Risk Management Division - works to maintain the company's stability by finding the balance between assets (insurance premiums) and public savings - executive funds and pension funds in order to pay its obligations. During the last years, a number of data science techniques are being adopted by the division to improve its predictions. As part of this effort, a collaboration between the Hebrew University of Jerusalem takes place in the form of a practicum at our seminar in data analytics. (Clal Insurance Ltd, 2021).
Clal presented to us a very challenging research problem which **purposed to find the most influential economic indicators, that based on their changes, the company can forecast the future tend in the stock markets.**
The stock market is one of the most attractive investment choices, one of them is **S&P500.** S&P500 stock market tracks the performance of 500 large U.S. companies by reporting the risks and returns of those companies. The stock market is risky and is known as volatile and unstable, where its price can increase one day, but a decrease in the next period (A. Dingli, 2017). The stock market performance represents the economic status of the country, as the fluctuation in its prices is caused by different factors such as: macroeconomic, microeconomic, social, political, psychological, etc. Therefore, the investors are always confused about the right time to invest in the stock market, and they always seek ways to have in advance information and analysis in order to take investment decisions as an accurate forecasting of stock market trends may yield profits for them, Clal is one of them.
The main purpose of this research is to build a high accurate model that predicts the future 90 days changes in S&P500 stock index based on the most influential U.S. economic indicators using different statistical, machine learning, deep learning algorithms, and finally compare metrics of those models.

Literature Search

Macroeconomic Indicators

The impact of macroeconomic indicators on stock markets as well as on their indices has been emphasized in scientific literature recently and has become more relevant for the two recent decades. It is often argued that stock prices are determined by some fundamental macroeconomic variables. Therefore, macroeconomic variables can influence investment decisions and motivates many researchers to investigate the relation between stock market prices and macroeconomic variables.

The relation between macroeconomic indicators and stock prices is confirmed in the most academic works, although there is a lack of comprehensive assessment of causality and dependence of macroeconomic indicators and stock market in regard to the time and changing macroeconomic processes. That is why the model of the impact of macroeconomic indicators on the stock market index, which enables to reveal a complex assessment of causality and dependence of the relation between macroeconomic indicators and stock prices during the long and the short runs, becomes a logical prolongation of an existent academic analysis (Pilinkus, 2010).

The analysis of the conception of macroeconomic indicators, principles of their classification and their place in the general system of economics is outlined in a number of scientific works (Bikker & Kennedy 1999; Rogers 1998; Mohr 1998; Darnay 1998; Dua 2004; Lakštutienė 2008; Tvaronavičius and Tvaronavičienė 2008, Chen 2009; Ciegis 2009). The relevance of macroeconomic indicators and possibilities of their use in the modern economic theories are emphasized by some other authors (Blaug 1997; Backhouse 2002; Tsai and Lee 2006; Norvaišienė, Stankevičienė, and Krušinskas 2008; Rutkauskas, Miečinskiene, and Stasytyte 2008; Snieška, Laskienė, and Pekarskienė 2008 ; Dumludag 2009).

Macroeconomic indicators are treated as statistical indicators which are used for assessment of general state of the country's economy during a certain period of time (Rogers, 1998) or as regularly published governmental statistics which reflects the economic situation in the specified country (Mohr, 1998). Macroeconomic indicators may be classified by their connection with the country's business cycle, the rate of declaration in different statistical editions, the character of economic process what facilitates initiative identification of certain economic processes. Chen et al. (1986) investigated whether a development of macroeconomic variables are risks that are rewarded in the stock market. The paper confirmed that interest rates, inflation or industrial production systematically affect stock market. Gjerde and Saettem (1999) examine the causal relation between stock developments and the macroeconomic variables. The results show a positive relationship for oil price index and real economic activity. Similarly, the study by Flannery and Protopapadakis (2002) revaluates the effect of some macro announcement series on US stock returns. Among these series, six macro variables, namely, balance of trade, housing starts, employment, consumer price index, M1 and producer price index seem to affect stock returns. On the other hand, two popular measures of aggregate economic activity (real GNP and industrial production) do not appear to be related with stock market development (Pražák 2018).

One of the most important and researched macroeconomic indicators is the interest rate. The interest rate can be an indicator in influencing the people's decision to spend or deposit the money and the business world's decisions to make loans for various purposes. When an interest rate is lower, more firms are willing to borrow money to expand their businesses, and the results are an increase in stock prices (Huang, Mollick, & Nguyen, 2016). Saymeh and Orabi (2013) used the Granger causality statistic to determine inflation causes interest rates, although each variable is an independent variable. The interest rates in monetary policy influenced the stock prices (Apergis & Eleftheriou, 2002; Assefa, Esqueda, & Mollick, 2017). The study of Tursoy (2019) shows a significant relationship between stock prices and domestic interest rates; interest rates have negatively affected the stock prices. However, Apergis and Eleftheriou (2002) show that interest rate correlates positively and insignificantly to stock prices. M. Reza Pahlawan (2020) however argues in his article that Exchange rates can affect domestic investments such as stocks. This situation will cause a decrease in demand for shares so that stock prices decline. It is also supported through previous research who claimed that stock prices and currency exchange rates have a positive correlation.

Is it possible to predict changes in the stock market and what is the way to do it?

Stock Market Prediction (SMP) is not a simple task due to its non-linear, dynamic, stochastic, and unreliable nature. The stock market is dependent on various parameters, such as the market value of a share, the company's performance, government policies, the country's Gross Domestic Product (GDP), the inflation rate, natural calamities, and so on.

The Efficient Market Hypothesis explains that stock market costs are significantly determined by new information, and follow a random walk pattern, such that they cannot be predicted solely based on past information. This was a widely accepted theory in the past. With the advent of technology, researchers demonstrated that stock market prices could be predicted to a certain extent. Historical market data, combined with the data extracted from social media platforms for example, can be analyzed to predict the changes in the economic and business sectors (Rouf ,2021).

Several methods can be used to predict a data set, one is the implementation of machine learning. Machine learning is classified as an artificial intelligence method and has been widely implemented in the tasks of classification, spam filtering, and forecasting. One machine learning method that can be used is an artificial neural network (ANN), one of the most accurate and widely used forecasting methods One development of the ANN model, recurrent neural network (RNN), is also one of the models that can be used to predict a time-series data pattern (M. Reza Pahlawan,2020).

Predicting Volatility in the S&P 500

Market volatility may seem difficult to analyze and predict due to its dependence on seemingly random, yet interdependent, variables. However, the aggregation of multiple economic variables, or indicators, can create accurate predictions when large amounts of economic data, the likes of which are abundant and continually gathered, are analyzed Machine learning algorithms can be used in conjunction with these indicators, whose data is drawn from large economic databases, in order to analyze and predict the volatility of the S&P 500 stock index (Kapoor, et al., 2017).

Prediction in times of crisis

A study conducted on the stock exchange in Indonesia (Sumarsan, 2020) tried to predict the changes during the COVID-19. The study successfully conducted stock price prediction and proved that prediction can also be made in times of crisis. In this study, the interest rates and exchange rates are the independent variables, and the stock market index is the dependent variable. The use of SPSS in this study varies from the other studies that use the Johansen cointegration test, generalized method of moments (GMM), and PLS-SEM. There are various methods to do forecasting, such as time-series forecasting, class of artificial intelligence models, different neural network models, frequency domain models (Mallikarjuna & Rao, 2019), and logistic regression models (Huet et al., 2004). The results showed no single forecasting technique provided uniformly optimal forecasting for all markets (Mallikarjuna & Rao, 2019). The forecasting of market fluctuation helps investors make the appropriate adjustment to the portfolios (Nguyen & Nguyen, 2019). Fourier transformations have been used in this study to forecast the financial time series. The study reviled four fresh and robust evidence. Partially, the interest rate has affected positively and significantly the stock market index. Partially, the exchange rate has affected negatively and significantly the stock market index. The F-test result, interest rate, and exchange rate have significantly affected the stock market index (JKSE) simultaneously. Furthermore, the FFT curve fitting has predicted that the stock market fluctuates and increases over time. The results have shown a strong influence of the independent variables and the dependent variable.

Research Methodology

This section will give a description of the used methodology and the performed tests in order to select the most influential economic indicators and to study if the changes in these variables have an effect on the volatility of the price of the S&P500 index. Both S&P 500 index close price and U.S economic indicators data are transformed to achieve stationarity and used in daily frequency. We used python programming language to execute this analysis on the period between 1986-12-31 / 2021-12-31. The research methodology steps are:

1) Data gathering and preprocessing.
2) Features selection.
3) Modelling and metrics.
4) Results comparison.

Data gathering and preprocessing

The data was collected using panda's data reader library from https://finance.yahoo.com websites to obtain the **closing price of the S&P 500 index** (dependent variable) and from https://fred.stlouisfed.org website to obtain **37 US economic indicators** that can potentially influence financial market performance (independent variables), these indicators are categorized in groups as shown in the **Table No.1**: Types of the Most Influential Indicators Affecting Stock Markets , Check Section 1:

Test and achieve Stationarity:

To find the best model to predict the future trend of the stock market, we used time-series analysis to find feasible information in our over time growing data. We applied it to the S&P 500 Stock Index and the economic factors. Time series data is a set of observations that's gathered over time, it is different from traditional predictive modeling techniques. When modeling, there are assumptions that we should take in account, such as stationarity. Time series data is stationary when the observations are not dependent on the time, thus the data will have constant mean, variance, and covariance. It is easier to model time series data when it is stationary. Moreover, most of the forecasting methods require stationarity to achieve effective predictions. Trend, seasonality, and other time-dependent structures can easily violate this assumption. Therefore, first we should check whether there is any evidence of a trend or seasonal effects and, if there is, to remove them. To this end, some methods are used[1] Check Section 3:

1) **Plots**: we can see in **Graph No.1** that S&P500 index has increasing trend and in **Graph No.2** there are clear patterns in June and November. More visualization for the economic indicators, Check Section 2.
2) **Summary Statistics**: by splitting the time series into two contiguous sequences, then calculating the mean and variance of each group. If the means and the variances are volatile, then the data is non-stationary.
3) **ADF Test:** If p-value > 0.05: fail to reject the null hypothesis that the data is stationary.

```
mean1=1004.646594
mean2=3626.961693
variance1=331248.414389
variance2=4082019.242430
```

```
ADF Statistic: 5.039421
p-value: 1.000000
Critical Values:
        1%: -3.431
        5%: -2.862
        10%: -2.567
```

From the previous results, we can confirm that our dependent variable (S&P500 index) is **non-stationary**. By repeating the earlier methods, we can check the stationarity of our independent variables (the economic indicators). As the time series is always growing, including non-stationary variables in the phase of modeling will give us bad results, where any model will learn on small levels during the first years and won't know what to predict when the input data is too high due to it never seeing those high levels during the training phase. In order to achieve stationarity, growth **transformations** on all non-stationary time series are applied by taking the difference between each value and its previous value such as day-over-day(dod), week-over-week(wow), month-over-month(mom), quarter-over-quarter(qoq), and year-over-year(yoy) growth rates in percentage. **Graph No.3** shows S&P index after transformation. All the transformed data is joined together, and the original non-stationary time series data are removed from the dataset, Check Section 4 and Section 5.

Features selection

Correlation Analysis:

The relation between macroeconomic indicators and stock prices is confirmed in the most academic works, although there is a lack of comprehensive assessment of causality and dependence of macroeconomic indicators and stock market (Pilinkus, 2010). Since its difficult to infer causality by keeping all the economic indicators constant, except the one whose influence we want to study in the real time, the correlations between our dependent

[1] Jason Brownlee on December 30, 2016, How to Check if Time Series Data is Stationary with Python (machinelearningmastery.com)

variable (SPX_growth_365d, SPX_growth_90d, SPX_growth_30d) and the independent variables after transformations were calculated to select which indicators have the most impact on the stock price. From the correlation analysis Check Section 6a, we can see that there is strong correlation between some of the economic factors and the growth of the stock market index (-0.6 to -0.4 to+0.54-0.62). Also, mostly the same factors highly correlated with SPX_growth (30d, 90d, and 365d), which proves the validity of our results. Despite of the fineness of the results, they are simultaneous and looking at the recent changes in the macroeconomic indicators against the current growth of the index, not forward-looking growth, therefore, it's not really helpful if we want to predict the future trend of the stock market. So, our next step is to find the correlations of the recent changes of the same macro indicators with the future changes in the stock market. Accordingly, we did another differencing transformation between each close value of S&P500 index and its value at a latter period, in this way we obtained SPX_future_ growth (30d, 90d, and 365d). From **Table No.2**, we can see the most (positively and negatively) correlated indicators with the 90-days future growth of S&P500. We can notice that the values of the correlations went down Check Section 6b (from +-0.4 for 365 days to +-0.15 for 30 days) when we tried to predict the growth for the latter period, which is logical as stock markets are volatile in the short term and hard to predict. Moreover, the impact of the changes in the macro factors on the stock market price is not immediate but grows over time.

Multicollinearity

The most correlated indicators with 90-days future growth of S&P that we found in the previous step gave us a starting point to select the features that we want to include in our forecasting model, but it's not enough because these features can be correlated with each other, as a result, we can't identify what causes the change of stock price in the first place. Therefore, our next step is to check the multicollinearity between them by using **VIF Test** and to get the features' importance by **Decision Tree method** to in order to find and rank the marginal impact of each factor Check Section 6c and Section 6d. VIF test values should be between 5 and 10, If the value is higher than 10, that indicates high correlation, and it's better to remove this feature because it leads to poorly estimation. From **Table No.3** we can see the listed features are after removing the indicators with the highest VIF.

According to the decision tree method, we can see in **Table No.4** the most important features when predicting the 90-days future growth for S&P500, they may not be highly correlated with the future growth of our index but can have an important marginal impact. Finally, we joined the most correlated features and the most important features together. VIF test was recalculated. The final results led us to the listed most influential indicators as in **Table No.5**

Modeling and metrics

We didn't model the future 30 days changes of the stock market because it's hard to predict as the stock market is chaotic in the short term. Other than that, general macroeconomic indicators provide a statistically significant impact on stock prices in the long run (Prazak's, 2018). Neither we didn't model the future 365 days change in the index as the time is money and the investors need to take short term decisions. Hence in this stage, we chose to model the future 90 days change of the stock market price based on the most influential **U.S. economic** indicators that we obtained in the previous stage. We used the supervised learning regression approach to analyze the current problem. From statistics we used liner regression. From machine learning we used Decision tree, and Prophet. From deep neural network algorithms, we used LSTM. The first step of this stage is to normalize the variables by using StandardScaler from sklearn library due to the different scales of our variables. The second step is to divide the dataset to training and testing sets, to fit our model to the training set, to make predictions on the test set, plot graphs of the results. The final step is to compare the metrics of the implemented models.

Linear Regression is used to investigate the relationship between the independent variables and the dependent variable to predict subsequent mean values of the variables. The coefficients below indicate whether the dependent variable increases or decreases as the regressors increase. The equation for our linear regression statistical model can be written as:

```
SPX_future_growth_90d = β0+β1*M1V +β2*DTWEXAFEGS +β3*T5YIFR +β54*SPCS20RSA+

β5*GFDEBTN vov+β6*PSAVERT+β7*CIVPART vov+β8*CP vov+β9*VIXCLS+ε
```

In our further implementation of this model, we found that the transformation that we did to achieve stationarity canceled the linear relationship between the dependent and independent variables, and no relationship was found, Check Section 7c

Decision Tree as mentioned previously, used as a baseline model to get the importance features, this model is simple to understand and to interpret. **Graph No.4** shows that the result of this model is not precise and doesn't really catch the actual trend and fluctuations in the price Check Section 7a

Prophet designed by Facebook, it is a time series forecasting library that tries to capture the seasonality in the past data and works well when the dataset is large. Actually, this model doesn't need any preprocessing in the data, but as we did a transformation to achieve stationarity, we built the model using the transformed data. **Graph No.5** shows the result of our model, we can see that the model could catch some trend of our target variable. For full implementation and visualization please Check Section 7d.

LSTM: After splitting the dataset into train and test sets, we split the train and test sets into input and output variables. Finally, the inputs are reshaped into the 3D format expected by LSTMs, namely [samples: 6986, timesteps: 14, features: 9]. We defined the LSTM with 50 neurons in the first hidden layer and 1 neuron in the output layer for predicting the change in the index. The input shape will be 14-time step with 9 features. We used the mean_squared_error (MSE) and the efficient Adam version of stochastic gradient descent. The model will be fit for 100 training epochs with a batch size of 32. **Graph No.6** shows the result of the model.

Evaluation of the models: After the model is fit, we can forecast for the entire test dataset. We combine the forecast with the test dataset and invert the scaling. With forecasts and actual values in their original scale, we can then calculate an error score for the implemented models. In this case, we calculated the Root Mean Squared Error (RMSE) Mean Absolute Error (MAE) and Mean Squared Error (MSE) as listed in **Table No. 6** Models Accuracy Metrics

Results

In this research, it was found that the future change in S&P 500 index is most significantly affected by 9 indicators as listed in **Table No.5**. The recent changes of them are negatively/ positively correlated with our target variable. Based on the recent changes of those indicators and by calculating the accuracy metrics, LSTM model gave us the best performance comparing to the other models. Although the model is not very precise, therefore it is not very safe to predict the future stock market movement well only by using macroeconomic indicators. As we saw earlier in the analysis, the economic factors gave limited explanation power. So, any conclusions regarding to the impact of the economic features on stock markets and how much they help to predict the trend of stock markets should be taken carefully.

Conclusions

Precise and consistent stock price prediction is a difficult task due to its noisy and nonlinear behavior. There are several factors that can impact the prediction such as fundamental market data, macroeconomic data, technical indicators, political conditions, and others. We developed different models to predict the future change of S&P 500 index's closing price by extracting a well-balanced combination of input variables, capturing the multiple aspects of the economy. The performance of the suggested models was analyzed by using various evaluation metrics to identify the best model. The experimental results show that LSTM model with around 50 hidden neurons can provide a superior fit and high prediction accuracy compared to other models.
The proposed model can be easily customized to apply in other broad market indexes where the data exhibits a similar behavior. Interested invest companies can use the proposed model to better inform the market situation before making their investment decisions. Our model can contribute to better understanding of the changes of direction of the stock market, and to expand it into the world of risk management and scenario analysis.

Future Directions

We managed in the existing model to find the changes of direction in 90 days; we would like to use more tests and analysis tools to find more factors that can catch the direction of changes a little earlier.
As for extreme scenarios (crisis), except the Covid period which did not come from the economic worlds then the expectation to catch something is low, but for real economic crises this model helps us base the claim that macroeconomics and markets are a bit disconnected in the last 12 years, based on Avi Ben-Noon, Chief Risk Officer in Clal Insurance.
The ability of macroeconomic factors to regulate markets which is largely the key parameter influencing market performance in recent years and how much central bank liquidity is flowing or not to markets. Therefore, it is a key parameter in decision making as well, and we would like to see in eventually, a model that will give us something about the intensity of economic events because it is a parameter that greatly affects when looking at portfolio, composition, venture capital and loss pads that investment firms hold.

In the future, we can explore the possibility of incorporating unstructured textual information in the model such as investor's sentiment from social media, earning reports of underlying companies, the immediate policy-related news, and research reports from market analysts.

Another potential direction of the future work can be developing hybrid predictive models by combining the LSTM with some other neural networks architectures. To improve the prediction accuracy even further, we also can implement hybrid optimization algorithms to train the model parameters by combining the existing local optimizers with the global optimizers such as genetic algorithms and particle swarm optimization algorithms (Lee & Kang, 2020).

Limitations

The stock price fluctuations are uncertain, and there are many interconnected reasons behind the scenes for such behavior. The possible cause could be the global economic data, changes in the unemployment rate, monetary policies of influencing countries, immigration policies, natural disasters, public health conditions, and several others.

With additional time, we could analysis more the feature selection from the financial data by combining some strategies together for instance to train the LSTM model by input of supervised classification model since the stock markets are naturally noisy, non-parametric, non-linear, and deterministic chaotic systems and to examine for which approach we get the most accurate results.

Another limit we faced is that the model's performance handles stationary time series data only by default, and performance can be totally inaccurate if it is used for non-stationary data. Therefore, it was essential for us to convert non-stationary time series data to stationary before implementation, which may lose the original structure and interpretability of the features.

With allocating budget for this project, we could purchase special consulting services to collect expanded data which will improve our results, some research focuses on complex statistical or machine learning techniques without focusing on the type of attributable variables. Others use only the fundamental data without exploring additional factors that could influence the stock market prediction. There is a need to develop a model with a good combination of features of the stock market variables and simplicity in model architecture. Thus, our contribution is to create a model without adding any complexity in model architecture and maintaining a well-balanced set of variables to capture the behavior of the stock market from multiple dimensions, but of course with outsider consulters our model will be much accurate.

Bibliography

Apergis, N., & Eleftheriou, S. (2000). Interest Rates, Inflation, and Stock Prices: The Case of the Athens Stock Exchange. *Journal of Policy Modeling*, 231-236.

Assefa, T. A., Esqueda, O. A., & Mollick, A. V. (2017). Stock Returns and Interest Rates Around the World: A Panel Data Approach. *Journal of Economics and Business*, 20-35.

BDI. (2021). *Ranking of insurance companies in Israel.* Retrieved from https://www.bdicode.co.il/category/heb_insurance/

Bikker, J. A., & Kennedy, N. O. (1999). Composite leading indicators of underlying inflation for seven EU countries. 225-258.

Blaug, M. (1997). *Economic Theory in Retrospect.* Cambridge University Press.

Chen, S. S. (2009). Predicting the bear stock market: macroeconomic variables as leading indicators. *Journal of Banking & Finance*, 211-223.

Ciegis, R. (2009). Theoretical Reasoning of the Use of Indicators and Indices for Sustainable Development Assessment . *Inzinerine Ekonomika – Engineering Economics*, 33-40.

Clal Insurance Ltd. (2021). *Analysts' presentation, fourth quarter.* Retrieved from https://www.clalbit.co.il/aboutclalinsurance/financialstatementsandpressrelease/

Darnay, A. J. (1998). *Economic Indicators Handbook: Time Series, Conversions, Documentation.* Detroit Gale.

Dua, P. (2004). *Business Cycles and Economic Growth : an Analysis Using Leading Indicators.* Oxford University.

Dumludag, D. (2009). An analysis of the determinants of foreign direct investment in Turkey: The role of the institutional context. *Journal of Business Economics and Management*, 15-30.

Flannery, M. J., & Protopapadakis, A. (2002). University of Southern California, Finance and Business Economics Department.

Gjerde, Ø., & Sættem, F. (1999). Causal relations among stock returns and macroeconomic variables in a small, open economy. *Journal of International Financial Markets, Institutions and Money*, 61-74.

Huang, W., Mollick, A. V., & Nguyen, K. H. (2016). U.S. Stock Markets and The Role of Real Interest Rates. *The Quarterly Review of Economics and Finance*, 231-242.

Huet, S., Bouvier, A., Poursat, M., & Jolivet, E. (2004). *Statistical Tools for Nonlinear Regression.* New York: Springer.

Kapoor, V., Khedkar, N., O'Keefe, J., Qiao, I., Venkatesan, S., & Laghate, S. (2017). Predicting Volatility in the S&P 500 through Regression of Economic Indicators. *New Jersey Governor's School of Engineering and Technology*.

Lakštutienė, A. (2008). Correlation of the indicators of the financial system and gross domestic product in European Union countries. *Inzinerine Ekonomika – Engineering Economics*, 7-18.

Lee, J., & Kang, J. (2020). Effectively training neural networks for stock index prediction: Predicting the S&P 500 index without using its index data. *Plos One*.

Mallikarjuna, M., & Rao, R. P. (2019). Evaluation of Forecasting Methods from Selected Stock Market Returns. *Financial Innovation*, 1-16.

Mohr, P. (1998). *Economic Indicators.* Pretoria : University of South Africa.

Nguyen, C. T., & Nguyen, M. H. (2019). Modelling Stock Price Volatility: Empirical Evidence from the Ho Chi Ming City Stock Exchange in Vietnam. *The Journal of Asian Finance, Economics and Business*, 19-26.

Norvaišienė, R., Stankevičienė, J., & Krušinskas, R. (2008). The Impact of Loan Capital on the Baltic Listed Companies' Investment and Growth. *Inzinerine Ekonomika – Engineering* , 40-48.

Pilinkus, D. (2010). Macroeconomic Indicators and Their Impact on Stock Market Performance in the Short and Long Run: The Case of the Baltic States. *Technological and Economic Development of Economy*, 291-304.

Pražák, T. (2018). The Effect of Economic Factors on Performance of the Stock Market in the Czech Republic. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 1613-1626.

Rogers, R. (1998). *Handbook of Key Economic Indicators.*

Rouf, N. (2021). *Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions* . MDPI.

Rutkauskas, A. V., Miečinskiene, A., & Stasytyte, V. (2008). *Investment decisions modelling along sustainable development concept on financial markets.* Technological and Economic Development of Economy.

Saymeh, A. A., & Orabi, M. M. (2013). The Effect of Interest Rate, Inflation Rate, GDP, on Real Economic Growth Rate in Jordan. *Asian Economic and Financial Review*, 341-354.

Snieška, V., Laskienė, D., & Pekarskienė, I. (2008). Stock returns and the macroeconomic environment: The case of the Vilnius stock exchange. *Transformations in Business & Economics*, 115-129.

Sumarsan, T. (2020). Determinants and Prediction of the Stock Market during COVID-19: Evidence from Indonesia. *The Journal of Asian Finance, Economics and Business*.

Tsai, M. T., & Lee, K. (2006). A study of knowledge internalization: from the perspective of learning cycle theory. *Journal of Knowledge Management*, 57-71.

Tursoy, T. (2019). The Interaction between Stock Prices and Interest Rates in Turkey: Empirical Evidence from ARDL Bounds Test Cointegration. *Financial Innovation*, 5-7.

Tvaronavičius, V., & Tvaronavičiene, M. (2008). Role of fixed investments in economic growth of country: Lithuania in European context. *Journal of Business Economics and Management*, 57-64.

A. Dingli, K. S. Fournier, "Financial time series forecasting - a deep learning approach," Int. J. Mach. Learn.

Comput., vol. 7, no. 5, pp. 118-122, 2017, doi: 10.18178/ijmlc.2017.7.5.632.

Appendices

Table No.1: Types of the Most Influential Indicators Affecting Stock Markets	
Category	**Indicators**
	GDP: Gross Domestic Product

Growth	GPDI: Gross Private Domestic Investment
Prices and Inflation:	CPIAUCSL: Consumer Price Index for All Urban Consumers: All Items in U.S. City Average GDPDEF: Gross Domestic Product: Implicit Price Deflator
Money Supply:	M1SL: M1 M1V: Velocity of M1 Money Stock M2V: Velocity of M2 Money Stock
Interest Rates	T5YIE: 5-Year Breakeven Inflation Rate T10YIE: 10-Year Breakeven Inflation Rate T5YIFR: 5-Year, 5-Year Forward Inflation Expectation Rate TEDRATE: TED Spread (DISCONTINUED)
Unemployment	UNRATE: Unemployment Rate CIVPART: Labor Force Participation Rate EMRATIO: Employment-Population Ratio UNEMPLOY: Unemployment Level MANEMP: All Employees, Manufacturing
Income &Expenditure	CDSP: Consumer Debt Service Payments as a Percent of Disposable Personal Income MDSP: Mortgage Debt Service Payments as a Percent of Disposable Personal Income FODSP: Household Financial Obligations as a Percent of Disposable Personal Income PSAVERT: Personal Saving Rate RSXFS: Advance Retail Sales: Retail Trade
Government Debt	GFDEBTN: Federal Debt: Total Public Debt GFDEGDQ188S: Federal Debt: Total Public Debt as Percent of Gross Domestic Product
Alternative investments	DCOILWTICO: Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma GVZCLS: CBOE Gold ETF Volatility Index
Exchange Rates	DTWEXBGS: Nominal Broad U.S. Dollar Index DTWEXAFEGS: Nominal Advanced Foreign Economies U.S. Dollar Index
Other indicators	INDPRO: Industrial Production: Total Index HOUST: New Privately-Owned Housing Units Started: Total Units TCU: Capacity Utilization: Total Index VIXCLS: CBOE Volatility Index: VIX DIVIDEND: Net corporate dividend payments STLFSI2: St. Louis Fed Financial Stress Index (DISCONTINUED) CP: Corporate Profits After Tax MORTGAGE30US: 30-Year Fixed Rate Mortgage Average in the United States SPCS20RSA: S&P/Case-Shiller 20-City Composite Home Price Index

Table No.2	
Feature	**Corr**
DTWEXAFEGS	0.2011
SPCS20RSA	0.2056
GFDEBTN_yoy	0.2057
PSAVERT	0.2243
DTWEXBGS	0.2521
M1V	-0.2498
T10YIE	-0.2362
T5YIFR	-0.2353
FODSP	-0.2045
T5YIE	-0.1971

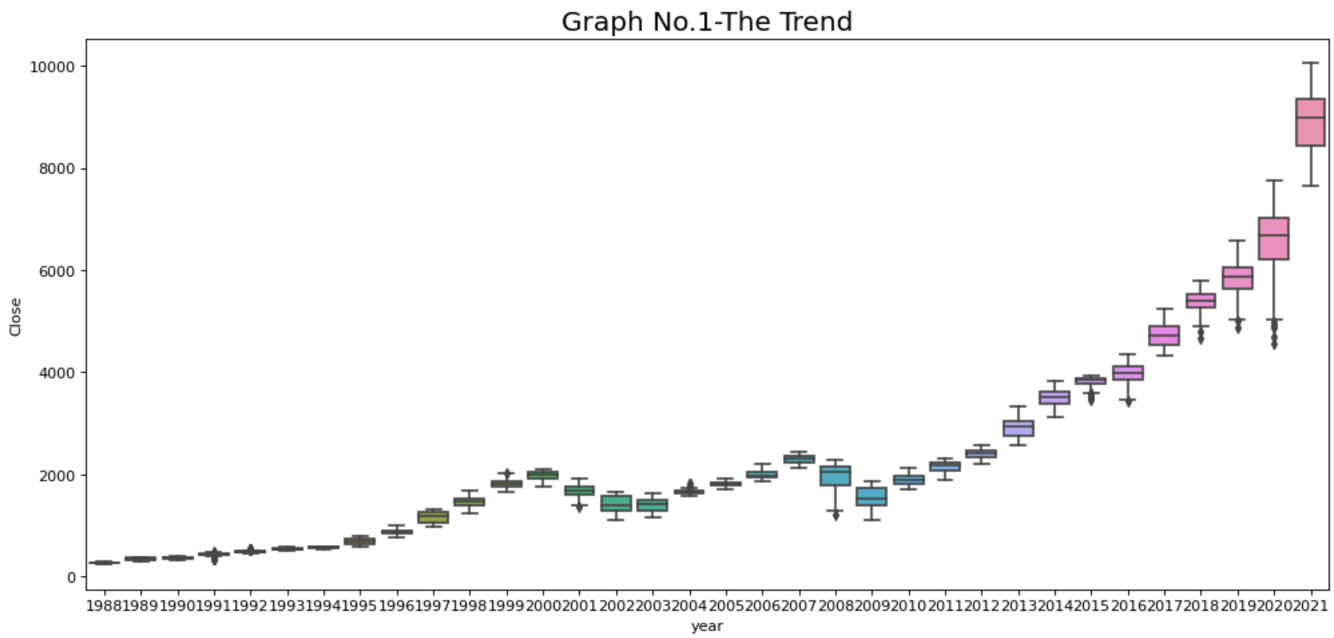
Table No.3	
Feature	**VIF**
M1V	4.586
DTWEXAFEGS	6.504
T5YIFR	10.54
SPCS20RSA	10.70
GFDEBTN_yoy	4.811
PSAVERT	6.101

Table No.4	
Feature	**Importance**
CIVPART_yoy	0.0462
FODSP	0.0547
CP_yoy	0.0630
VIXCLS	0.0749
div_ratio	0.0846
M1V	0.1377

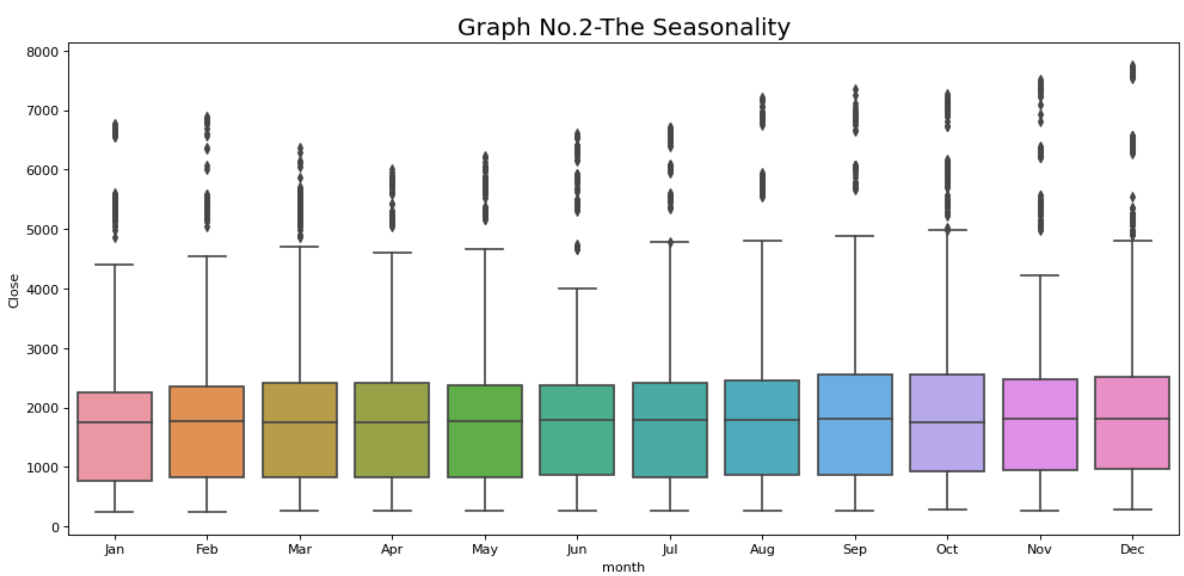
Table No.5	
Feature Symbol	**Feature Name**
M1V	Velocity of M1 Money Stock
DTWEXAFEGS	Nominal Advanced Foreign Economies U.S. Dollar Index
T5YIFR	5-Year Forward Inflation Expectation Rate
SPCS20RSA	S&P/Case-Shiller 20-City Composite Home Price Index
GFDEBTN_yoy	Federal Debt: Total Public Debt
PSAVERT	Personal Saving Rate
CIVPART_yoy	Labor Force Participation Rate
CP_yoy	Corporate Profits After Tax
VIXCLS	CBOE Volatility Index: VIX

Table No. 6 Models Accuracy Metrics			
Model/Metric	**MSE**	**RMSE**	**MAE**
Decision Tree	1%	10%	8%
Prophet	2%	14%	10%
LSTM	0.5%	7%	5%

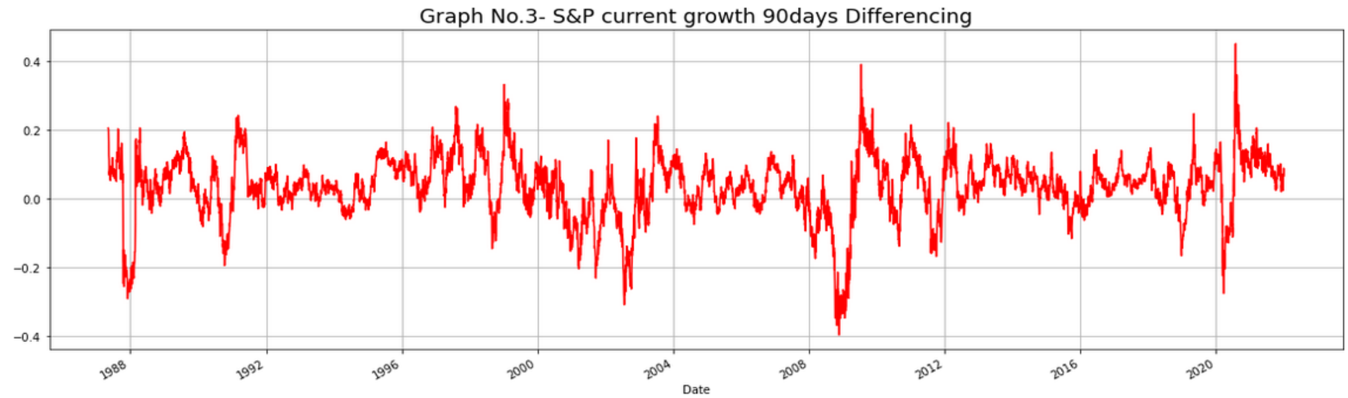
Graph No.1



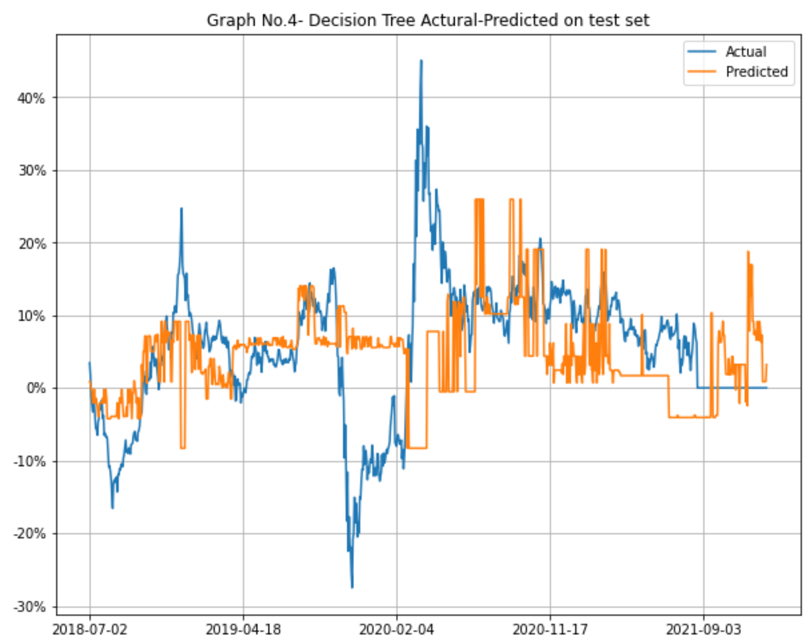
Graph No.2



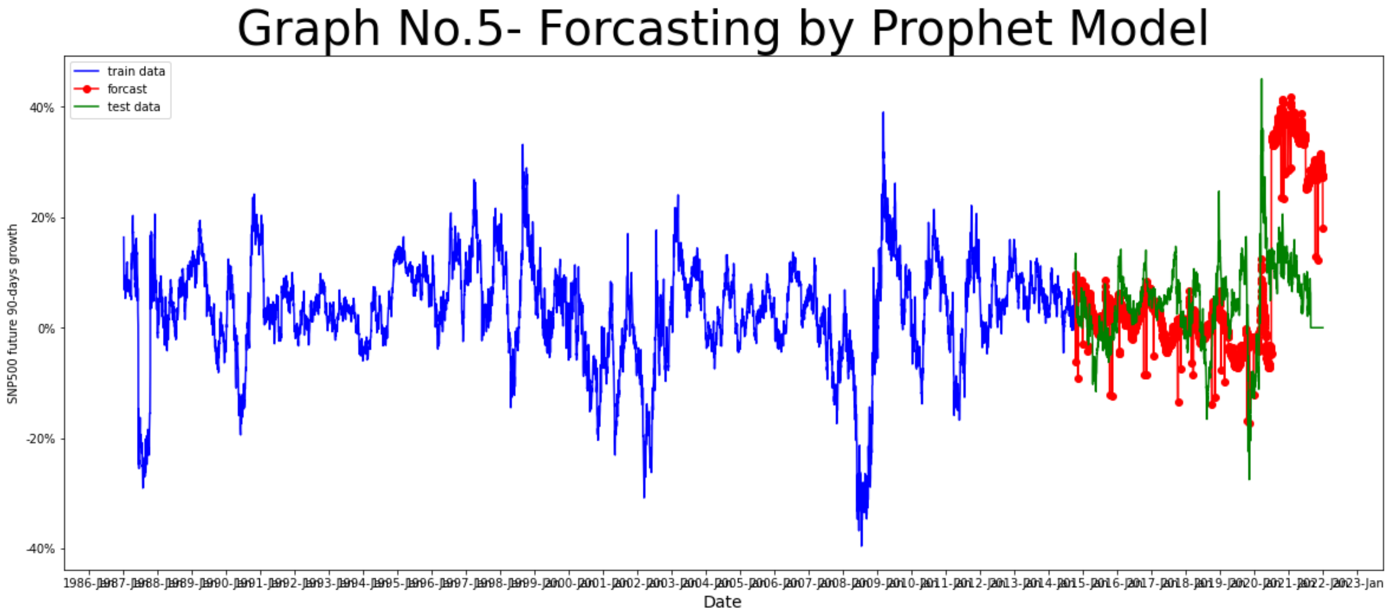
Graph No.3



Graph No.4



Graph No.5



Graph No.6