

Forecasting GDP based on Stock Market Indices

¹Suyog Mahale, ²Sujal Joshi, ³Sarang Phalak, ⁴Anshuman Dipanshu, ⁵Shubham Sharma

^{1,2,3,4,5}Student

^{1,2,3,4,5}Computer Engineering Department

^{1,2,3,4,5}D.Y Patil College of Engineering, Akurdi, Pune, India

Abstract - GDP is a fundamental measure of the economic development of a country. A big contributor to the GDP of any country are the financial markets of that country. In this paper, we aim to use popular stock market indices of India (SENSEX, NIFTY50) to predict the change in GDP of the subsequent financial quarter. Previous researches on GDP prediction incorporate financial market data into their models however they do not particularly focus on the effects of the financial market features. Our model focuses on the effects financial markets have on the GDP of India. We will be creating a machine learning model utilising the Support Vector Regression module in the Support Vector Machine algorithm (SVM).

Index Terms - Gross Domestic Product (GDP), Machine Learning (ML), Preprocessing, Support Vector Machine (SVM), Support Vector Regression (SVR), Features, Elastic Net Regression, Shapley Value

I. INTRODUCTION

Gross Domestic Product (GDP) is a quantitative measure of all financial goods and services sold within a country over a particular period of time. GDP is the sum of Consumer Spendings (C), Investments (I), Government Spendings (G), and the difference between Exports and Imports (X-M).

$$GDP = C + I + G + (X - M)$$

We can determine the economic state of a country based on its GDP and change in GDP over time. GDP estimates in India are released by the Ministry of Statistics and Programme Implementation. We get an estimate of the change in GDP over a financial quarter. Properly forecasting and predicting GDP is imperative for the financial stability of a country. With proper estimates we can prepare for disaster or plan development ahead of time. One major contributing factor to GDP are the financial markets of a country. In India, we have two major stock indices. NIFTY50, which represents the weighted average of 50 of the largest companies listed on the National Stock Exchange. SENSEX, a free-float market-weighted, stock market index of 30 well-established companies on the Bombay Stock Exchange (BSE). By analysing the change in these two stock market indices, we can create a model which predicts the change in GDP of India for the next quarter.

II. LITERATURE SURVEY

The research paper, Is Accounting Useful for Forecasting GDP Growth? (Srikant Datar, Apurv Jain, Charles C.Y Wang, Siyu Zhang et al. 2021), is a comprehensive insight onto the effects that accounting features have on GDP prediction. The researchers have taken the past 120 quarters of data of GDP estimates made by The Bureau of Economic Analysis (BEA) and The Survey of Professional Forecasters (SPF). They have also taken 341 features out of which 292 features are related to accounting. The researchers used Elastic Net Regression to train their model and evaluated the model using the Holdout MSE algorithm. Finally they calculated the Shapley Values for each of the features that they incorporated into their model to determine the importance of each feature.

The model created by the researchers had approximately 28 features related to the stock markets. However, these features were not individually analysed for their direct effect on GDP prediction. Taking inspiration from this paper, we have decided to focus our research on the effect that changes in stock markets can have on the GDP of a country.

III. DATA COLLECTION AND PROCESSING

For our model, we have used the past 15 years, monthly data of NIFTY50 and SENSEX indices. We will be using the High, Low, Open, Close, and Change features from the indice data. We have used the past 15 years quarterly data of GDP change and estimates from the Ministry of Statistics and Programme Implementation.

Since the data for GDP is quarterly, we have to convert our data from NIFTY50 and SENSEX into similar time frames. Each financial year has four quarters starting from June.

Q1: June-August

Q2: September-November

Q3: December-February

Q4: March-May

We will have to convert the NIFTY50 and SENSEX data into a similar quarterwise format. The open price is calculated from the month at the start of the quarter while the close price will be from the month at the end of the quarter. The high and low points will also be calculated between each quarter. We will be performing our data processing operations on MS Excel. We have split the data into 80% (12 Years) data for training and 20% (3 Years) data for testing.

Due to the scarcity of training data, we have combined both the NIFTY50 and SENSEX datasets while training our model. The final model will produce results based on the general change between either of the stock market indices.

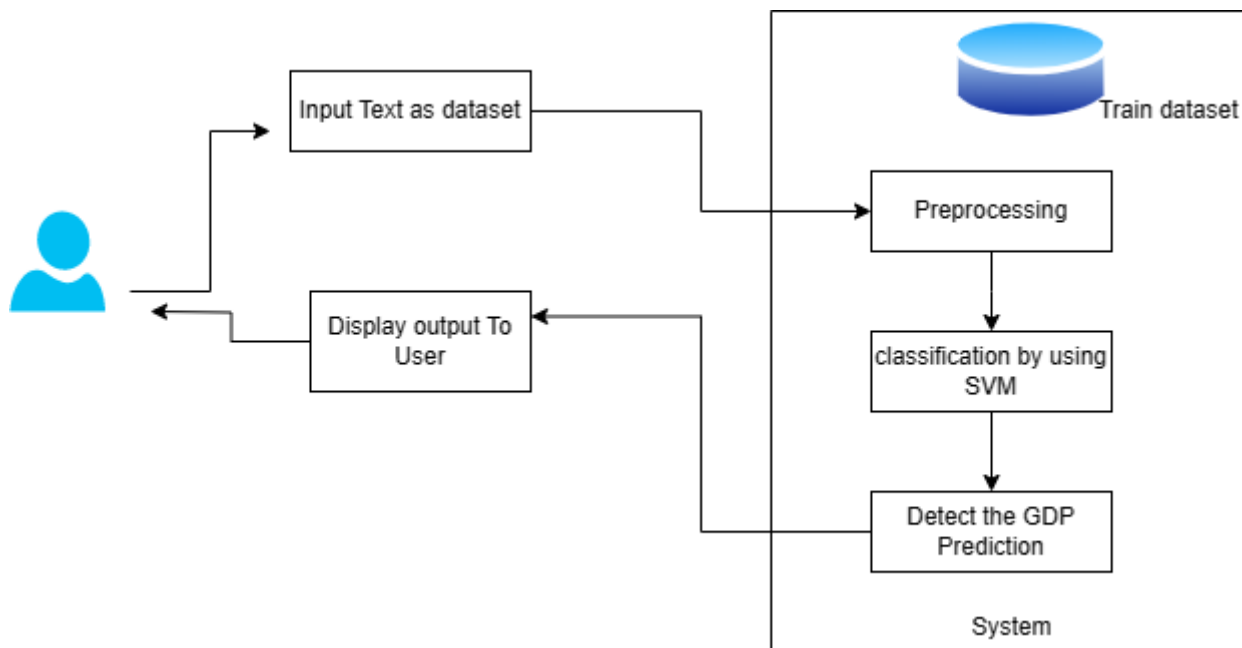


Figure 1. Model Architecture

IV.METHODOLOGY

4.1 Anaconda - Anaconda is an open source coding and development platform for data science applications and is a must use for building and deploying data science models. We will be using the spyder application in the anaconda platform. Spyder is an open source development environment for scientific programming in the python language. All our code will be written on the spyder platform and executed as an application.

4.2 Support Vector Machine(SVM) - Support Vector Machine is a supervised learning algorithm which is used for both classification and regression problems. SVM algorithm creates vectors of each of our input parameters and accordingly creates a hyperplane which best represents or classifies these data points. For regression We will be using the regression instance, Support Vector Regression(SVR), for training our model.

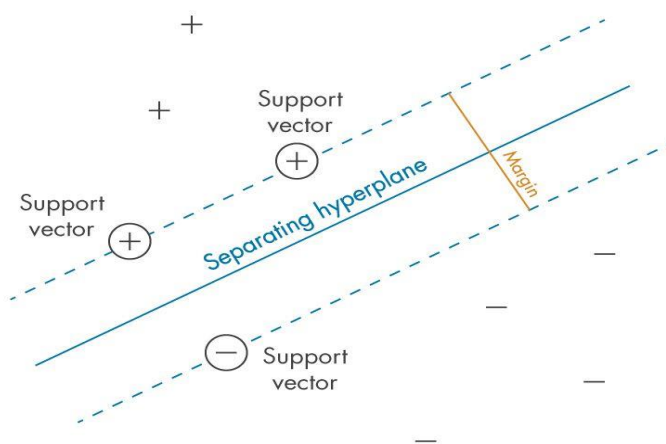


Figure 2. Support Vector Machine Logic

4.3 Support Vector Regression(SVR) - SVR instance can be imported from the SVM library using `from sklearn.svm import SVR`. We will be specifying the kernel, C, gamma, and epsilon inputs for the SVR function. We will be leaving the rest of the parameters on their default values.

We will be using a linear kernel. In SVR, regression is performed at a higher dimension as there are multiple features in our data. The kernel helps map these data points into the higher dimension. The selection of the kernel determines the behaviour of the model therefore it is important to select a suitable kernel for our regression problem.

C is the regularisation parameter in our model. It defines the penalty term to be introduced while calculating our estimated data points. C is inversely proportional to the penalty term being introduced. We are setting the value to C to 100 for our prediction model.

The Gamma term defines how far the influence of one point reaches within the rest of the model. A higher value of gamma implies that each point has more influence along the rest of the model. We will be setting the value of gamma to be 'auto'.

The value of epsilon determines the width of the tube around our hyperplane. Only points within this tube are considered while mapping the target variable. A lower value of epsilon indicated a smaller margin of error. We will be setting the value of epsilon to 0.1 in our model.

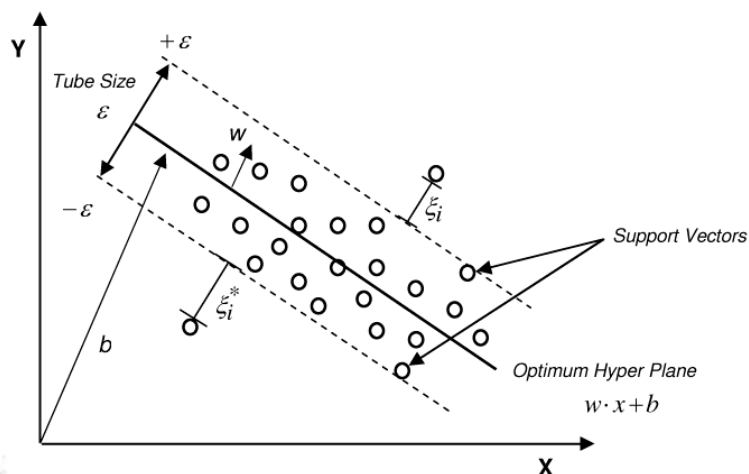


Figure 3. SVM for Linear Regression

4.4 Joblib - Joblib is a python library which is used for saving and loading machine learning models for faster parallel execution. In joblib, data is stored as byte strings rather than objects. This allows for faster and more compact data storage. Joblib has two instances in it, dump and load. Dump is used to store a model in the system and load is used to load the model in our code. We can use joblib by first installing the package into our environment and using `from joblib import dump, load` to load joblib functions into our code.

4.5 Process - To create our model first we define a function ModelTraining() in our code. We will import our dataset into the function. There is no need to process this data as we have already preprocessed this data in MS Excel. Next step is to define our SVR model along with all the parameters(kernel, C, epsilon, gamma). We will fit our model with the training dataset containing x_train and y_train. We will use the model on the testing data, x_test, to predict our target variable which is the actual change in GDP for the next quarter. Once we are done building our model we will save it as a joblib file in the folder. The file is saved using the 'dump' instance of the joblib library. We can call this joblib file during execution.

The next step is to create a variable for each of our input features. We create a function Prediction() and map each of our input features to a variable. Next, we import the joblib file which we have saved in our folder. We will use our model to predict the target variable by using the variables which we have created as input to the model. Finally, we will display the output to the user.

V. RESULTS

We have given the input as Open, High, Low, Close, and percent Change metrics of NIFTY50 or SENSEX over a financial quarter along with Previous GDP change and next quarter GDP forecasts. The SVR algorithm maps these features and gives us an estimate of how much percent change in GDP we can expect for the next quarter.

We found that our predicted value of GDP for our testing set, deviated from the actual values by an average of 18.46%. This comes out to about 0.012 to 0.029 deviation from the actual values. While this value may not seem significant, it has larger implications when considering a large value such as the GDP of the country. Another attributing factor to the high deviation is the scarcity of data.



Figure 4. GDP Prediction Main page

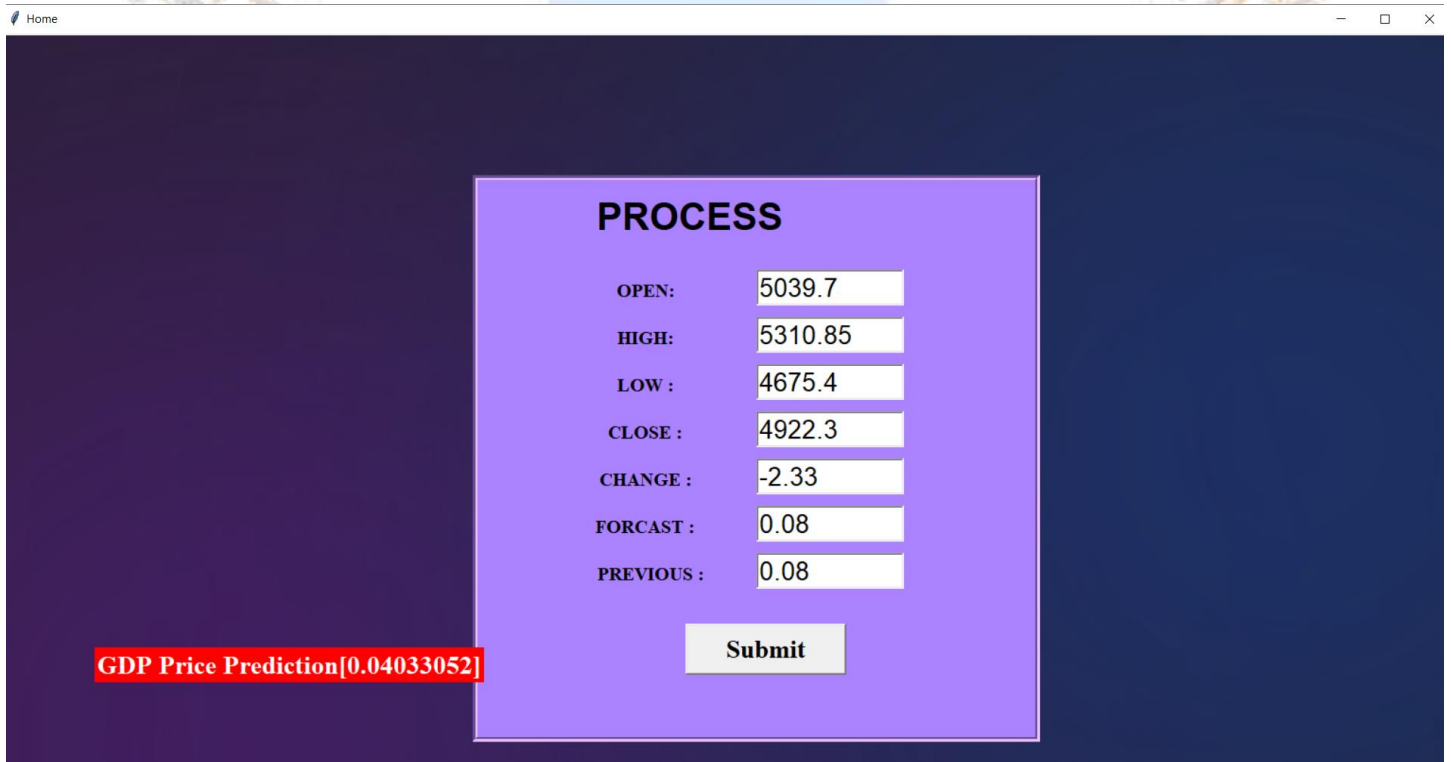


Figure 5. GDP Prediction process

VI. CONCLUSION

We have successfully created a demonstration of the effect that stock market fluctuations in India have on the GDP of the country. We have used the Support Vector Regression algorithm to create our model. From our results we can conclude that changes in the stock market can give us a rough estimate of the direction of change in GDP. However, there is not enough data to produce an accurate model related to only stock market features. This model, in combination with other features related to GDP prediction, can give us a fair idea of the GDP estimate for further quarters. Financial investors can accordingly manage their investments and the country can make preparations and plans for the further development of the country.

VII. REFERENCES

- [1] Srikant Datar, Apurv Jain, Charles C.Y. Wang, Siyu Zhang - Is Accounting Useful for Forecasting GDP Growth? A Machine Learning Perspective
- [2] Umme Salma, Md. Mahmudul Hasan, Mst Irin Sultana - A Regression Model Based Approach to Identifying Determining Factors for GDP Growth in Bangladesh
- [3] Sandeep Kumar Amit K. Shukla Pranab K. Muhuri Q.M. Danish Lohani - Transfer Learning based GDP Prediction from Uncertain Carbon Emission Data
- [4] Sandeep Kumar Amit K. Shukla Pranab K. Muhuri - Isolation Forest Based Multi -Source Unsupervised Transfer Learning for Missing GDP Prediction.
- [5] Amit K. Shukla Sandeep Kumar Rishi Jagdev Pranab K. Muhuri Q. M. Danish Lohani - Interval Type-2 Fuzzy weighted Extreme Learning Machine for GDP Prediction.
- [6] K. Amadeo, "What Is the GDP Growth Rate?" The Balance, 28 March 2019.
- [7] M. Jahn, "Artificial neural network regression models: Predicting GDP growth," Hamburg Institute of International Economics (HWWI), Hamburg, 2018.
- [8] Figure 2 from in.mathworks.com
- [9] Figure 3 from researchgate.net

