Predicting Macroeconomic Trends using Predictive Modelling in R

Devang Manjramkar, Dhananjay Deore, Ayush Gundawar, Ishan Gujarathi

Department of Engineering, Sciences and Humanities (DESH)

Vishwakarma Institute of Technology, Pune - 411037, Maharashtra, India

***Abstract –*** Macroeconomy in **easy** **phrases** **manner** the **exact** **examine** of **overall performance** and **conduct** of the economy. Macroeconomics analyzes all **mixture** **signs** and the **elements** that **have an effect on** the economy. In our project, **we've** **attempted** **to apply** Data Science and **numerous** **strategies** **to be had** in Data Science **inclusive of** Data Visualization, Data Analysis, Regression **version** **just like the** Linear Regression Model, Model Performance on **those** macroeconomic datasets **and feature** **attempted** to **are expecting** the **destiny** **traits** and **apprehend** **the connection** **among** the variables. With the **assist** of R-studio as IDE and availability of **a couple of** in-**constructed** **programs** and libraries, **we've** **attempted** **to plan** **numerous** graphs **among** macroeconomic dataset variables, **evaluate** and **apprehend** **the connection** **among** variables and their **effect** on **destiny** prediction of **traits**. We have then **carried out** a Regression **version** i.e the Linear Regression Model **at the** **testing** dataset **and feature** **attempted** to **are expecting** the macroeconomic **traits** **and feature** **concurrently** analyzed the **version** **overall performance** and accuracy **in conjunction with** it.

***Keywords – Machine Learning, Regression, Linear Regression, Exploratory Data Analysis, Macroeconomic Trends, Predictive Modelling***

***­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­***

I. INTRODUCTION

The **fast** **improvements** in **information** and **communication** **technology** **during the last** **many years** have produced an explosive **increase** in **the quantity** of **records** collected, **main** to **the brand new**

**technology** of **massive** **statistics**. The **records** has **caused** the **improvement** **of latest** **technology** in **diverse** fields of statistics, **gadget** **studying** and **statistics** mining, **additionally** interacting with the fields of engineering and **synthetic** intelligence (AI), **amongst** others. This **big** **attempt** **caused** the **beginning** of **the brand new** interdisciplinary **area** called "Data Science", whose **standards** and **strategies** **intention** **on the** **automated** extraction of **probably** **beneficial** **records** and **know-how** from **statistics**. Although **statistics** **technological know-how** **technology** **were** **efficiently** **carried out** **in lots of** **special** fields (e.g., healthcare, predictive maintenance, and **deliver** chain management), their **capability** has been little explored in **monetary** **statistics**., **monetary** and macroeconomic. In this project, we **attempted** **to use** **statistics** **technological know-how** and its **special** **strategies** like visualization, regression, etc. on **diverse** macroeconomic **statistics** **units** to **expect** **diverse** parameters.

Macroeconomic **statistics** **particularly** **includes** **mixture** values of **monetary** flows **each** at **the extent** of **the full** economy, **inclusive of** GDP and **countrywide** income. Using **diverse** **statistics** **technological know-how** **strategies** and **integrated** libraries **to be had** in R-Studio, we made predictions on **those** macroeconomic variables **the use of** forecast **fashions** **inclusive of** the linear regression **version** and **tried** to **expect** trends. futures. In our project, we **advanced** a **version** that explains the relationships **among** **those** factors. These macroeconomic **fashions**, and the forecasts they produce, are **utilized by** **authorities** **corporations** to **assist** **assemble** and **verify** **monetary**, monetary, and **monetary** policy; **through** **corporations** to **outline** the **approach** in **countrywide** and **international** markets; and **through** **traders** to forecast and plan **moves** in **diverse** asset classes.

II. METHODOLOGY

In this paper we have presented three different case studies of three different datasets i.e., Healthcare Finances, Bitcoin Price, House Prices which all are macroeconomic trends nowadays (bitcoin has occupied a larger space in lives of Indians and around the world therefore it can now be considered under macroeconomy). We have followed the recommended data science pipeline of data cleaning, exploratory data analysis, prediction using machine learning algorithms and then inference.

Through this pipeline we have successfully built 3 ML models (mainly as some others were rejected due to performance but kept part of the R code for reference) which can be used to predict these three respective macroeconomic trends.

We have also visualized the plots for EDA and regression outputs on every stage with various plotting techniques like ggplot2 and corrplot.

III. Algorithms Used

We have used the linear regression machine learning algorithm for predicting three macroeconomic trends namely house prices, bitcoin price, and healthcare finances. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable. The variable we are using to predict the other variable's value is called the independent variable. This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a “least squares” method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

IV. Tools and Technologies

In our Project, we have used R-Studios as the IDE and its various inbuilt packages and libraries. We have used the following libraries in our project which are available in R studios:

**Caret (Classification and Regression Training):** It contains the functions to streamline the model training process for complex regression and classification problems.

**Corrplot:** It provides visual exploratory tools for correlation amongst the variables and supports automatic variables reordering to help detect hidden patterns among the variables.

**FactoMineR:** It is used for multivariate exploratory data analysis and data mining.

**Ggplot2:** It is used for data visualization.

**dplyr:** It is used for making data manipulation easier.

**Hmisc:** It contains many functions useful for data analysis, high level graphics and utility operations.

**Cowplot:** It is a simple add-on to the ggplot2 library. It has features that help in creating quality figures such as a set of themes, functions to align plots etc.

**GGally:** It extends ggplot2 by adding several functions to reduce the complexity of geometric objects with transformed data.

**E1017:** It provides functions for statistical and probabilistic algorithms.

V. Results and Discussion

We have obtained accurate predictors on each stage while it was difficult to do so at the beginning due to low model performances, we still improved it throughout. We have done the data visualization beautifully for the world to get a quick and thorough glimpse into what the model is predicting. We found that all the steps of the data science pipeline are as crucial as the main ML techniques. Data cleaning, visualization, and analysis are very crucial in deploying any real-world solutions.

VI. Annexures

Chart

Description automatically generated with low confidence

Fig 1. Correlation Plot for Bitcoin

Chart, scatter chart

Description automatically generated

Fig 2. Correlation Plot for Healthcare Finance Dataset

Chart, bar chart, waterfall chart

Description automatically generated

Fig 3. Correlation Plot for House Price Prediction Dataset

Chart, scatter chart

Description automatically generated

Fig 4. Linear Regression Model

VII. Futurescope

We can improve the prediction of the model by using deep learning and reinforcement learning algorithms. Also, many advanced data visualization libraries are coming out now which can help visualize the data in 3D and in a cleaner way. Also, we can explore R shiny and presentations schemes for more deeper research projects like these for better presentation of the results.

VIII. Conclusion

We infer from this data science activity that we can very effectively predict crucial macroeconomics using a well-defined and curated data science pipeline. It is very crucial to use the correct tools, libraries, and functions in all the stages of this pipeline to get the most optimal output at the end of that stage. We have also inferred here that most of the macroeconomic trends can be predicted using simple machine learning algorithms like linear regression and polynomial regression (if required) and do not require sophisticated algorithms at least on reasonably small datasets. Big data analysis might benefit from these advanced algorithms and even deep learning though.

IX. References

1. Robin F. Niesert, Jochem A. Oorschot, Christian P. Veldhuisen, Kester Brons, Rutger-Jan Lange, Can Google search data help predict macroeconomic series? International Journal of Forecasting, Volume 36, Issue 3, 2020, Pages 1163-1172, ISSN 0169-2070, https://doi.org/10.1016/j.ijforecast.2018.12.006

2. Bin Weng, Waldyn Martinez, Yao-Te Tsai, Chen Li, Lin Lu, James R. Barth, Fadel M. Megahed, Macroeconomic indicators alone can predict the monthly closing price of major U.S. indices: Insights from artificial intelligence, time-series analysis and hybrid models, Applied Soft Computing, Volume 71, 2018, Pages 685-697, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2018.07.024.

3. Monica Lam, Neural network techniques for financial performance prediction: integrating fundamental and technical analysis, Decision Support Systems, Volume 37, Issue 4, 2004, Pages 567-581, ISSN 0167-9236, https://doi.org/10.1016/S0167-9236(03)00088-5

4. Perron P. (1994) Trend, Unit Root and Structural Change in Macroeconomic Time Series. In: Rao B.B. (eds) Cointegration. Palgrave Macmillan, London. https://doi.org/10.1007/978-1-349-23529-2\_4

5. M. Chudý and E. Reschenhofer, “Macroeconomic Forecasting with Factor-Augmented Adjusted Band Regression,” Econometrics, vol. 7, no. 4, p. 46, Dec. 2019.

6. Smith, T. A., & Rajan, A. (2017). A Regression Model to Predict Stock Market Mega Movements and/or Volatility Using Both Macroeconomic Indicators & Fed Bank Variables. International Journal of Mathematics Trends and Technology, 49(3). <https://doi.org/10.14445/22315373/IJMTT-V49P522>

7. Hongxiang Sun, Zhongkai Yao, Qingchun Miao, "Design of Macroeconomic Growth Prediction Algorithm Based on Data Mining", *Mobile Information Systems*, vol. 2021, Article ID 2472373, 8 pages, 2021. <https://doi.org/10.1155/2021/2472373>

8. Aksoy, Yunus, Henrique S. Basso, Ron P. Smith, and Tobias Grasl. 2019. "Demographic Structure and Macroeconomic Trends." *American Economic Journal: Macroeconomics*, 11 (1): 193-222.DOI: 10.1257/mac.20170114

9. Bitcoin Price Prediction and Analysis Using Deep Learning Models.DO - 10.1007/978-981-15-5397-4\_63

10. Olatunji, I. A., Wahab, B. M., Ajayi, M. T. A., & Liman, H. S. (2016). Influence of macroeconomic factors on residential property returns in Abuja Nigeria. ATBU Journal of Environmental Technology, 10(1), 67–83.

11. Anghelache, C., G.R. Pagliacci, M., Prodan, L., (2013). "Model for macroeconomic- analyze based on the regression function", Romanian Statistical Review, no. 1/2013, ISSN 1018-046X.

12."[**Models and indicators used in macroeconomic forecast**](https://ideas.repec.org/a/rsr/supplm/v65y2017i3p40-48.html)," [Romanian Statistical Review Supplement](https://ideas.repec.org/s/rsr/supplm.html), Romanian Statistical Review, vol. 65(3), pages 40-48, March.

13. M. A. and Ismail, S. (2012). Multiple Regression in Analyzing House Price Variation. IBIMA Journal.2012, article ID

14 Y. Pan, Z. Xiao, X. Wang, and D. Yang, “A multiple support vector machine approach to stock index forecasting with mixed frequency sampling,” *Knowledge-Based Systems*, vol. 122, pp. 90–102, 2017.View at: [Publisher Site](https://doi.org/10.1016/j.knosys.2017.01.033) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=A%20multiple%20support%20vector%20machine%20approach%20to%20stock%20index%20forecasting%20with%20mixed%20frequency%20sampling&author=Y.%20Pan&author=Z.%20Xiao&author=X.%20Wang&author=&author=D.%20Yang&publication_year=2017)

15. Jiawei Han, MichelineKamber, “Data Mining Concepts and Techniques”, pp. 279-328, 2001.