TOPOLOGICAL APPROACHES TO GDP PREDICTION USING TDA

*Dr.T.A.Albinaa
*Assistant Professor, PSGR Krishnammal College for Women, Coimbatore

Abstract:
The topic of GDP has become of high importance among the indicators of economic variables. GDP prediction is a crucial job in the economy and growth analysis of a country. The goal of the paper is to give a different approach concerning the classical econometric techniques, and to show how Machine Learning techniques may improve calculating the Gross Domestic Product accurately. The GDP of countries is impacted by various social, economic, and cultural parameters. We have analysed those parameters utilizing the Kaggle Dataset, persistent diagram were used in the present study to forecasts GDP of a country demonstrating the application of Topological Data Analysis.

Key Words: Gross Domestic Product, Topological data analysis, Persistent diagram, Jupyter Notebooks.

Introduction:

GDP is most extensively used to estimate the economy's output or production. It is termed as the total value of goods and services produced within a country in a specific time period. GDP is an precise indicator of the size of an economy and the GDP growth rate indicates the economic growth. GDP per capita refers to the correlation with the trend in living standards over time. The national income and product accounts (NIPA) forms the basis for measuring GDP. It helps the policymakers and central banks to judge whether the economy is contracting or expanding which enables to take necessary action.GDP can be calculated either through the expenditure, income, or value-added approach.

GDP calculation can be done as an expenditure approach or income approach i.e., (the total sum that everyone in an economy spent over a particular period or total sum earned in an economy by everyone). Another method is the value added approach which calculates the GDP by industry. Expenditure based approach is most common and produces both real and nominal values rather income based approach is done by nominal values. Samuelson and Nordhaus neatly sum up the importance of the national accounts and GDP in their seminal textbook "*Economics*." They liken the ability of GDP to give an overall picture of the state of the economy to that of a satellite in space that can survey the weather across an entire continent.

GDP enables policymakers and central banks to judge whether the economy is contracting or expanding, whether it needs a boost or needs to be restrained, and if threats such as a recession or rampant inflation loom on the horizon. The national income and product accounts (NIPA), which form the basis for measuring GDP, allow policymakers, economists, and businesses to analyze the impact of such variables as monetary and fiscal policy, economic shocks, such as a spike in the oil price, and tax and spending plans on specific subsets of an economy, as well as on the overall economy itself.Along with better-informed policies and institutions, national accounts have contributed to a significant reduction in the severity of business cycles since the end of World War II.

GDP can be calculated either through the expenditure approach—the sum total of what everyone in an economy spent over a particular period—or the income approach—the total of what everyone earned. Both should produce the same result. A third method, the value-added approach, is used to calculate GDP by industry. Expenditure-based GDP produces both real (inflation-adjusted) and nominal values, while the calculation of income-based GDP is only carried out in nominal values. The expenditure approach is the more common one and is obtained by summing up total consumption, government spending, investment, and net exports.

$GDP = C + I + G + (X – M)$

An ample of research has been performed on GDP among those many papers, several works have been overviewed and noted. Robert Costanza, Maureen Hart, Stephen Posner and John Talberth have worked on, GDP : The need for new measures of Progress in 2009. In 2012, Spencer L James, Paul Gubbins, Christopher JL Murray and Emmanuela Gakidou gave the notion of Developing a comprehensive Time Series of GDP per capita for 210 countries from 1950 to 2015. Impact of GDP, Spending on R&D, Number of Universities and Scientific Journals on Research Publications among Asian Countries was analyzed by Meo SA in 2013. Dr. V. Rama Devi gave an overview of Study on Predictors of GDP: Early Signals in 2014. Martin Feldstein has done his research in Underestimating the Real Growth of GDP, Personal Income, and Productivity in 2017. An Econometric Time Series GDP Model Analysis: Statistical Evidences and Investigations was studied by Habib Ahmed Elsayir in 2018. Arvind Subramanian initiated his work on India’s GDP Mis-estimation: Likelihood, Magnitudes, Mechanisms and Implications in 2019. In 2020, THW analyzed Japan’s Productivity and GDP Growth: The Role of Private, Public and Foreign R&D 1967–2017. Jan P. A. M. Jacobs, Samad Sarferaz, Jan-Egbert Sturm & Simon van Norden - Can GDP Measurement Be Further Improved? Data Revision and Reconciliation in 2020. Examining GDP Growth and Its Volatility: An Episodic Approach was researched by Jakub Bartak, Lukasz Jablonski and Agnieszka Jastrzebska in 2021.

An overview of the existing research work shows difference in opinions and ideas of Researchers based on GDP related to the global development. In order to sought this problem, this paper aims to compute the role of GDP as an approach in Topological Data Analysis. To perform this analysis, data is been collected and processed to acquire the topological structure. In other words, Topological Data Analysis reveals that data has shape and as we extract the topological features from the data collected it leads to several relationships and patterns. GDP termed as Gross Domestic Product determines whether an economy experiences growth or recession. The real time data being collected, persistence diagram is computed for the data and the relation between GDP, health and wealth of the population is been obtained

Proposed Methodology:

Dataset Collection:

The GDP Dataset is taken from Kaggle data to predict factors influencing the growth of GDP. The dataset consists of 260 countries with 20 different parameters. The parameters that are taken in consideration while predicting GDP are Literacy, Net migration, Population, Infant mortality, Agricultural economy, Industrial economy, Services economy, etc.

Topological Data Analysis:

TDA is a new area of study in complex data analysis that examines some unique geometric qualities known as "topological properties" that can be true even when a graph's shape is constantly changing. The topological characteristics of data in each spatial dimension can be accurately calculated using algebraic operations. For example, in a two-dimensional space, the primary elements include the number of points and their connectivity, while in a three-dimensional space, the main elements consist of the number of hollow spheres and their connectivity.

Vietoris Rips Complex:

The initial data shape is expressed by TDA using a straightforward complicated approach. One or more simplexes make up a simple complex, that are easier to work with computationally and mathematically than the original visuals and can simulate more complex shapes.

Persistent Homology:

One technique for calculating a space's topological properties at various spatial resolutions is persistent homology. Detected across a broad range of spatial scales, more persistent features are thought to be more likely to reflect genuine characteristics of the underlying space than sampling artifacts, noise, or specific parameter selection choices. Persistent homology typically takes a point cloud or function as input, and produces a persistence graph or landscape as output, depending on the type of analysis.

Data Interpretation

The Persistent homology is used for data projections into a Persistent diagram. It is built using Jupyter Notebook[16]. Persistent Diagram can be represented in the birth – death plane or the birth – lifetime plane. The birth – death plane is the representation pair(x,y) where x is the time of birth and y is the time of death of the feature in the persistence diagram.

Data Pre-processing and Cleaning:

Data pre-processing is required for cleaning the data and making it suitable. It helps to increase the accuracy and efficiency of a model. Identifying and removing errors and duplicate data, in order to create a reliable dataset is the main aim of data cleaning. For each dataset, we calculated the Euclidean distance matrix, which represents the pairwise distances between data points. Using the function (calculate_homology) computes the Vietoris-Rips complexes for each dataset. It represented by the distance matrix, helping to identify topological features like connected components and loops. These features are represented in the persistent diagram.

Interpretation of Persistent Diagram in Jupyter Notebook:

```
from ripser import ripser
from persim import plot_diagrams
```

```
distance_matrix = np.linalg.norm(gdp_data[:, np.newaxis] - gdp_data, axis=2)
```

```
diagram = ripser(distance_matrix, maxdim=2)['dgms']
```

```
plot_diagrams(diagram, show=True)
```

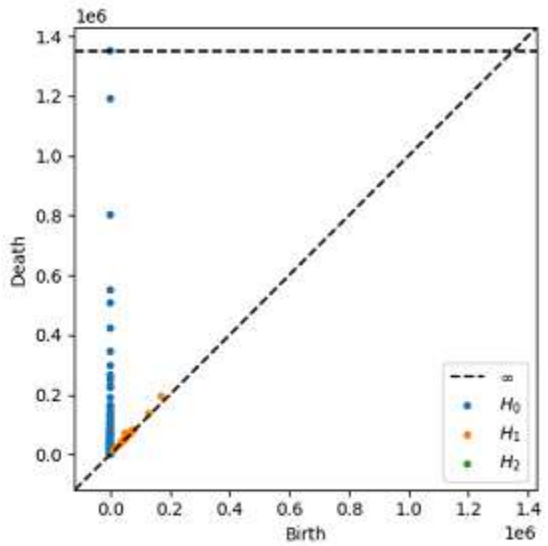

FIGURE 1. Persistent Diagram for GDP Dataset

Data Evaluation:

Random Forest is one of the well-known machine learning algorithms that belongs to the supervised learning technique category. Random Forest can be used both for regression and classification problems in machine learning. It is basically a classifier that consists of decision trees of the given dataset on numerous subsets. Further, the algorithm takes the average in order to improve the forecasting accuracy. Predictions from each tree that is formed are taken into consideration instead of just relying on a single decision tree and after that based on majority votes of prediction, output is predicted. When we choose the features for the Random Forest Classifier, our suggested model performs noticeably better. During the assessment, the highest accuracy rate was obtained to be 87%.

Finally, we are able to show the vectorization of Persistent Diagrams in Machine Learning classifiers.

Conclusions:

In a developing country like India, evaluating economic metrics such as GDP plays an important part in taking decisions both in private sectors as well as public sectors. In this paper, Persistent Diagrams in Machine Learning algorithms was predicted. Finally deployed the highest accuracy model to "GDP Estimation Tool" which estimates and forecasts GDP of a country just by giving some attribute as input for that country..

References:

[1] Aguilar, Alejandro, and Katherine Ensor. "Topology data analysis using mean persistence landscapes in fi-nancial crashes." Journal of Mathematical Finance 10.4 (2020): 648-678.
[2] Bang, James & Sen, Tinni & Basuchoudhary, Atin. (2015). New Tools for Predicting Economic Growth Using Machine Learning: A Guide for Theory and Policy
[3]Branco, Sérgio, et al. "Persistence Landscapes—Implementing a Dataset Verification Method in Resource-Scarce Embedded Systems." Computers 12.6 (2023): 110.
[4]Bubenik, Peter. "Statistical topological data analysis using persistence landscapes." J. Mach. Learn. Res. 16.1 (2015): 77-102.
[5] Bubenik, Peter, and Paweł Dłotko. "A persistence landscapes toolbox for topological statistics." Journal of Symbolic Computation 78 (2017): 91-114.
[6]Chazal, Frédéric, and Bertrand Michel. "An introduction to topological data analysis: fundamental and practical aspects for data scientists." Frontiers in artificial intelligence 4 (2021): 667963.
[7]Chang, Victor, et al. "Pima Indians diabetes mellitus classification based on machine learning (ML) algo-rithms." Neural Computing and Applications 35.22 (2023): 16157-16173.
[8]Edelsbrunner, Herbert, and John L. Harer. Computational topology: an introduction. American Mathemat-ical Society, 2022.
[9]Economy of India, Wikipedia - https://en.wikipedia.org/wiki/Economy_of_India#cite_note-53
[10]Fasy, Brittany Terese, et al. "Introduction to the R package TDA." arXiv preprint arXiv:1411.1830 (2014).
[11] Nicolau, Monica, Arnold J. Levine, and Gunnar Carlsson. "Topology based data analysis identifies a sub-group of breast cancers with a unique mutational profile and excellent survival." Proceedings of the Na-tional Academy of Sciences 108.17 (2011): 7265-7270.
[12]Padmawar, Vaishnavi & Pawar, Pradnya & Karande, Akshit. (2021). Gross Domestic Product Prediction using Machine Learning. 08. 2395-0056
[13] Somasundaram, Eashwar V., et al. "Benchmarking r packages for calculation of persistent homology." The R journal 13.1 (2021): 184.
[14]Shultz, Christopher. "Applications of Topological Data Analysis in Economics." Available at SSRN 4378151 (2023).
[15] Singh, Yashbir, et al. "Topological data analysis in medical imaging: current state of the art." Insights into Imaging 14.1 (2023): 58.
[16] Ulf W.Liebal, Rafael Schimassek, “ Biological Data Analysis Training with Jupyter Notebooks”,J Microbial Biol Educ,24(1), 2023
[17]Yoon, J. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. Comput Econ 57, 247–265 (2021)