Report On

International Debt Analysis Using Big Data

Submitted in partial fulfillment of the requirements of the Course project in Semester VII of Final Year Artificial Intelligence and Data Science

by

Shubhamkar Patra (Roll No. 35)

Chetan Sapkal (Roll No. 37)

Supervisor

Prof. Bhavika Gharat

University of Mumbai

Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence and Data Science

(2023-24)

Vidyavardhini's College of Engineering & Technology
Department of Artificial Intelligence and Data Science

CERTIFICATE

This is to certify that the project entitled “International Debt Analysis Using Big Data” is a bonafide work of "Shubhamkar Patra (Roll No. 35), Chetan Sapkal (Roll No. 37)" submitted to the University of Mumbai in partial fulfillment of the requirement for the Course project in semester VII of Final Year Artificial Intelligence and Data Science engineering.

Supervisor
Prof. Bhavika Gharat

Dr. Tatwadarshi P. N.
Head of Department

Abstract

International debt analysis plays a crucial role in understanding the financial relationships between countries and their economic stability. This study focuses on leveraging big data to analyze bilateral debt data between India and six South Asian nations - Bangladesh, Bhutan, Sri Lanka, Maldives, Myanmar, and Nepal obtained through the World Bank API. The dataset includes four key columns: 'Year' indicating the year of debt, 'Debtor' representing the debtor country code, 'Debt in US$' detailing the amount of debt in US dollars, and 'YoY Growth %' indicating the year-on-year growth percentage of bilateral debt. Through this analysis, we aim to shed light on the dynamics of debt relationships in the region and identify trends that may impact the economic stability of these countries.

Table of Contents

Pg. No

Chapter # 1

1.1 Problem Statement:

The problem at hand is to conduct a comprehensive analysis of international debt data between India and six South Asian countries, utilizing big data techniques. The specific challenges include identifying patterns, trends, and correlations within the dataset understanding the factors driving debt accumulation, and assessing the impact of year-on-year changes in bilateral debt. This analysis can provide insights into the financial relationships between these countries, and the potential risks or opportunities they may present to their respective economies.

Chapter # 2

2.1 Description and Working:

The analysis of international debt using big data, particularly focusing on bilateral debt data between India and six South Asian nations (Bangladesh, Bhutan, Sri Lanka, Maldives, Myanmar, and Nepal) obtained through the World Bank API, involves a multifaceted process. The initial step revolves around obtaining historical debt data, such as 'Year,' 'Debtor,' 'Debt in US$,' and 'YoY Growth %,' via the World Bank API. This data forms the foundation for understanding the economic relationships between these countries. Raw data is rarely ready for analysis. Data preprocessing is essential to ensure its quality and consistency. It encompasses cleaning the data, handling missing values, and converting data types as necessary.

Predictive modeling enables forecasting of future debt trends and assessing their impact on the countries involved. Classification models can categorize nations based on debt trends, making way for targeted policy recommendations. Interactive dashboards using tools like Plotly or Bokeh facilitate intuitive data exploration, aiding policymakers and stakeholders in making informed decisions.

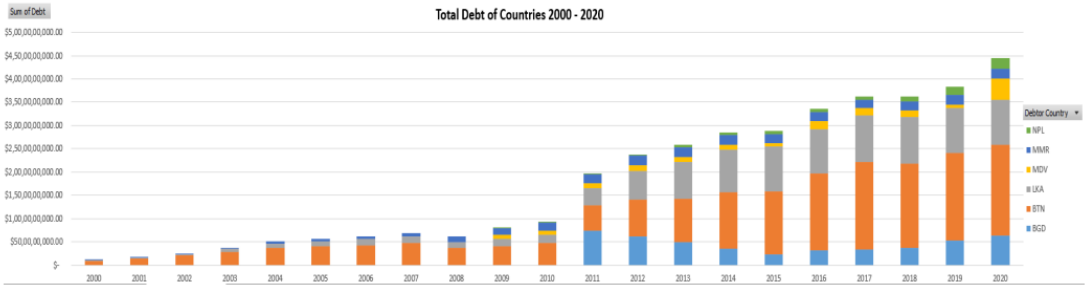


Fig. 1: India and Neighboring Countries PPG Bilateral Lending 2000-2020

2.2 Software & Hardware Used:

Software:

- Visual Studio Code
- Python 3.11
- Windows 10 OS
- Google Colab

Hardware:

- 64 bit Operating System
- 6gb RAM
- Intel i5 processor

Chapter # 3

3.1 Code:

In [1]:
```
# Import required packages
import requests
import json
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
import seaborn as sns
import matplotlib.pyplot as plt
```

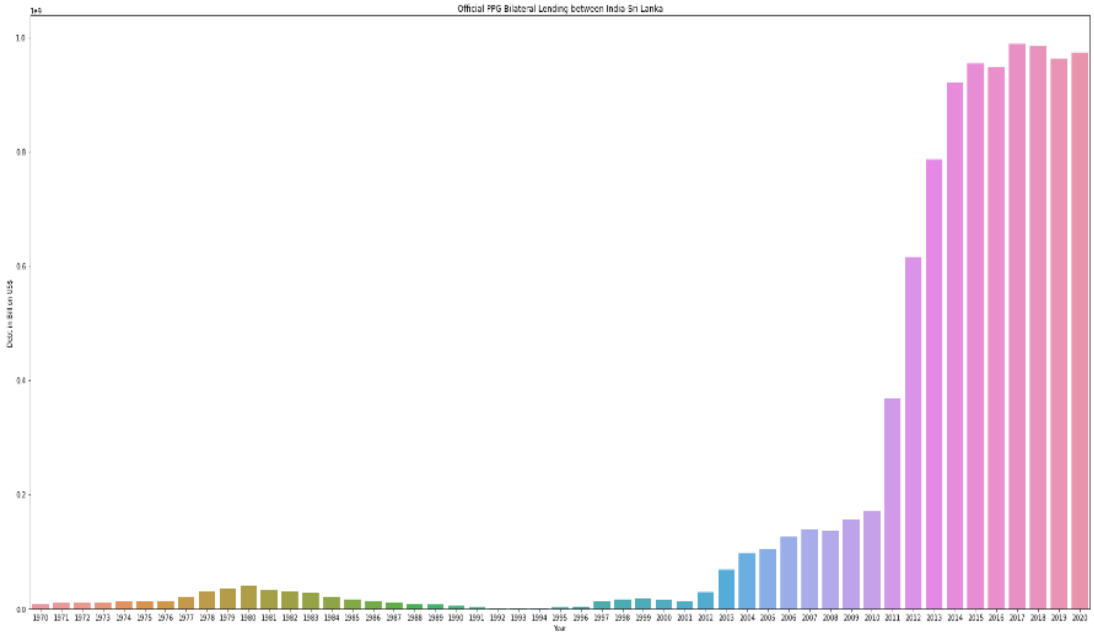
In [3]:
```
#Specifying Debtor Country & Creditor Country to check data
debtorCountry = input("Enter Debtor Country Code \n")
creditorCountry = input("Enter Creditor Country Code \n")
series = "DT.DOD.BLAT.CD"
time = "All"
```

In [4]:
```
# Setting up the API URL
url = "http://api.worldbank.org/v2/sources/6/country/"
end = "?format=json&per_page=500"
path = url+debtorCountry+"/series/"+series+"/counterpart-area/"+creditorCountry+"/time/"+time+end

# Creating a funtion that will parse through the JSON response
def getData(JSON):
    df = pd.DataFrame(columns=["year", "creditor", "debtor", "indicator","data"])
    for i in range(0,listLen):
        time = JSON["source"]["data"][i]["variable"][1]["value"]
        num = JSON["source"]["data"][i]["value"]
        df = df.append({"year":time, "creditor": creditorCountry,
                        "debtor":debtorCountry, "indicator":series, "data":num
                       }, ignore_index = True)
    return(df)
```

In [5]:
```
# Getting the data from the API
custom = requests.get(path)
customJSON = custom.json()
listLen = int(len(customJSON["source"]["data"]))
```

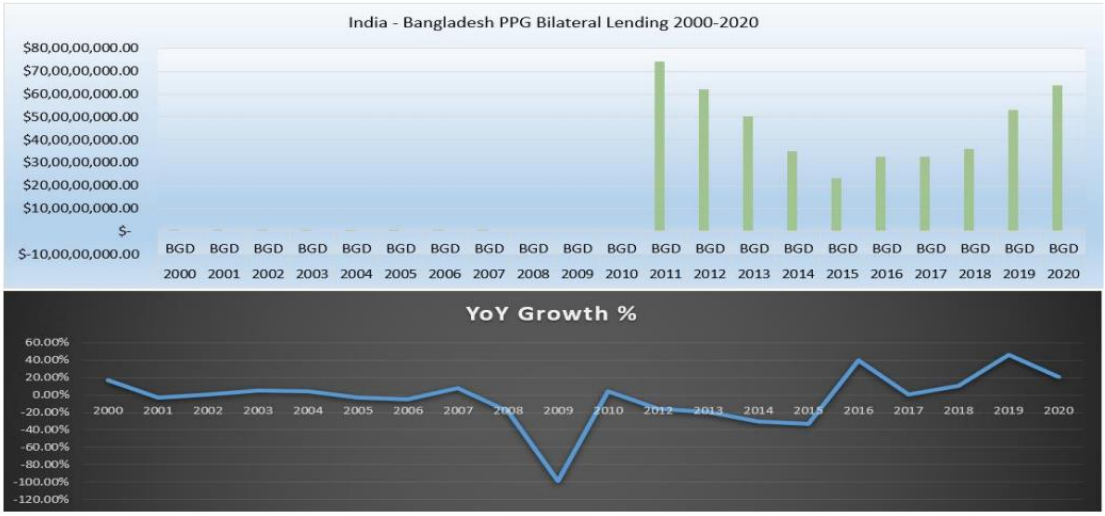
In [6]:
```
# Plugging the data into the parsing function and printing the data
IDSdata = getData(customJSON)
datatoexcel = pd.ExcelWriter(f'{debtorCountry}-{creditorCountry} PPG Bilateral Debt.xlsx')
IDSdata.to_excel(datatoexcel)
datatoexcel.save()
print("Excel File Saved")
```

In [16]:
```
# Selecting the dataframe created above as the data source for the chart
source = IDSdata.dropna()


# Plugging in the datasource, X and Y indicators, and the title for the chart
plt.figure(figsize=(30,15))
chart = sns.barplot(data = source, x= "year", y="data")
chart.set(xlabel='Year', ylabel='Debt in Billion US$', title='Official PPG Bilateral Lending between India-Sri Lanka')

# Displaying the chart
plt.show(chart)
```


Official PPG Bilateral Lending between India Sri Lanka
1e9
1.0
0.8
0.6
0.4
0.2
0.0
Debt in Bil on US$
1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
Year

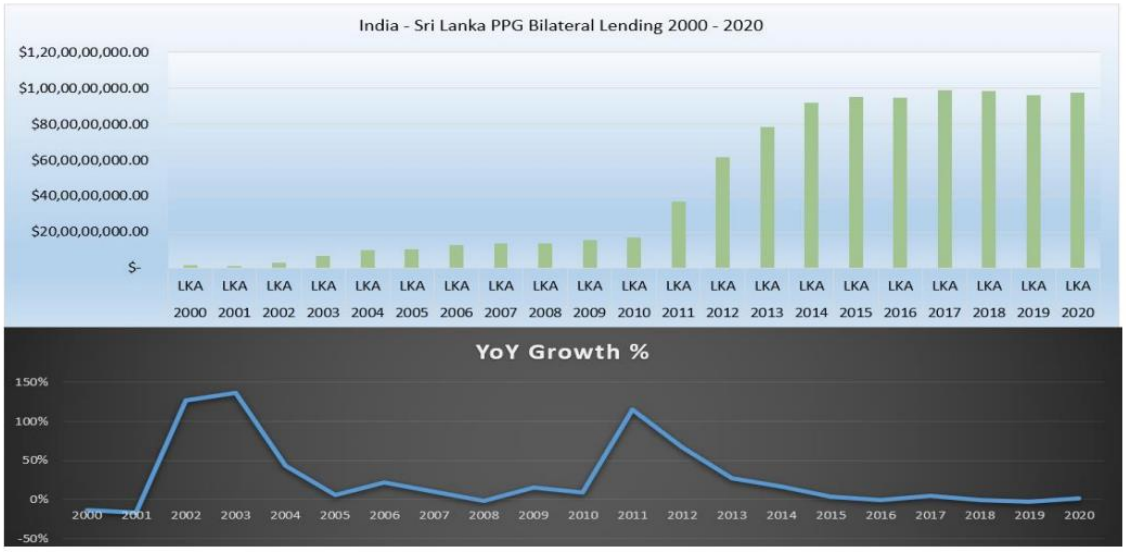
3.2 Results:

India – Bangladesh PPG Bilateral Lending 2000 - 2020

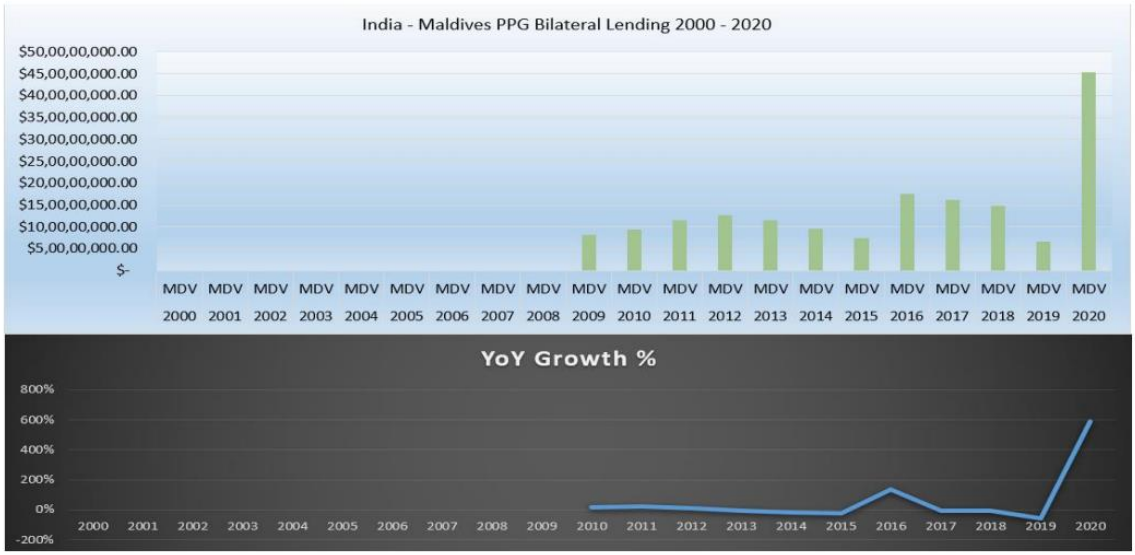


India – Bhutan PPG Bilateral Lending 2000 - 2020

India – Sri Lanka PPG Bilateral Lending 2000 - 2020



India – Maldives PPG Bilateral Lending 2000 - 2020



Conclusion and Future Work:

This international debt analysis using big data revealed valuable insights into the debt dynamics between India and several South Asian countries. Key findings include identifying trends in debt accumulation, understanding the factors driving these trends, and assessing the year-on-year growth of bilateral debt. Such analysis can inform policy decisions and guide financial strategies for the countries involved, promoting economic stability and cooperation in the region.

In the future, this research can be extended in several ways:

- Inclusion of more countries to provide a comprehensive regional perspective.
- Analysis of debt repayment patterns, including principal and interest payments.
- Integration of economic and geopolitical events to better understand debt fluctuations.
- Developing predictive models for early warning systems to manage debt crises.

Chapter # 4

References

[1] Kou, Gang, Yi Peng, and Guoxun Wang. ”Evaluation of clustering algorithms for financial risk analysis using MCDM methods.” Information Sciences 275 (2014): 1-12.

[2] Ding, Chris, and Xiaofeng He. ”K-means clustering via principal component analysis.” Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.

[3] Pham, Duc Truong, Stefan S. Dimov, and Chi D. Nguyen. ”Selection of K in K-means clustering.”Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 219.1 (2005): 103-119.

[4] Lloyd, Stuart P. ”Least squares quantization in PCM.” Information Theory, IEEE Transactions on 28.2 (1982): 129-137.

[5] A. Jain, R.C. Dubes,”Algorithms for clustering Data” in , Prentice Hall, 1988.

[6] Ester, Martin, et al.”A density-based algorithm for discovering clusters in large spatial databases with noise.” Kdd. Vol. 96. No. 34. 1996.

[7] Dudoit, Sandrine, and Jane Fridlyand. ”A prediction-based resampling method for estimating the number of clusters in a dataset.” Genome biology 3.7 (2002): research0036-1.

[8] Thalamuthu, Anbupalam, et al.”Evaluation and comparison of gene clustering methods in microarray analysis.” Bioinformatics 22.19 (2006): 2405-2412.

[9] Desgraupes, Bernard. ”Clustering indices.” University of Paris OuestLab ModalX 1 (2013): 34