|  |  |
| --- | --- |
| **Programme Title** | MSc in Business Analytics |
| **Module Title** | Data Analytics and Predictive Modelling |
| **Module Code** | 38157 |
| **Assignment Title** | Individual Proposal (40%) |
| **Level** | PG - Semester 1 (2023-2024) |

|  |
| --- |
| **Proposal Title:** |
| Decoding Development: Insights into Global Economies through Clustering and Predictive Modelling. |
| **Student ID** |
| 2709878 |

# 

# Abstract/Executive Summary:

In this proposal, we delve into crucial global economic indicators with an aim to pinpoint fundamental contributors to economic prosperity and stability. The highlight of our study is on GDP, GNI, sectoral output, and the balance of trade, with a focus on data from the most recent years.

Our dataset is comprehensive, covering a lot of economic indicators across a variety of nations, and is characterized by diverse data types and structures. We are adopting a unique analytical strategy, utilizing K-Means for clustering countries into meaningful groups, K-Nearest Neighbors and Decision Tress for sorting based on trade balance, and Linear Regression and Time Series Analysis to predict future economic patterns.

We expect this proposal to reveal key drivers of economic success, offer a closer look at how different nations withstand economic challenges, and shed light on how trade imbalances affect overall economic health. Ultimately, our goal is to find patterns and relationships that can provide strategic decision-making for governments, investment firms, and global agencies.

# Table of Contents

[Abstract/Executive Summary: 2](#_Toc149425595)

[Table of contents 3](#_Toc149425596)

[Introduction: 4](#_Toc149425597)

[Data gathering and data analysis: 6](#_Toc149425598)

[Feature Selection: 6](#_Toc149425599)

[Data Gathering: 9](#_Toc149425600)

[Data Analysis: 9](#_Toc149425601)

[Conclusion: 11](#_Toc149425602)

[Bibliography: 12](#_Toc149425603)

# Introduction:

The complex structure of global economies demands a good grasp of the elements that play a role in a country’s financial stability and capacity for expansion. Ensuring a nation maintains economic equilibrium and manages development is vital not just for its individual success, but also for the well-being of the global economy, given the interconnected nature of economies in the modern era.

**Business Challenge:**

Exploring and predicting the stability and development prospects of nations' economies is an important task. This not only aids governments in policy formulation but also assists investors in making strategic decisions, and aids international bodies in resource allocation.

**Significance:**

The significance of understanding and predicting a nation’s economic resilience and growth potential is multifold, especially in our globally interconnected economic landscape:

* **Investor Confidence**: Identifying stable economies instills confidence in investors, guiding them towards safe investment options.
* **Policy Formulation and Aid Allocation**: For governments and international bodies, this analysis is crucial for determining which regions may require economic assistance or intervention, and for devising policies that promote economic stability.
* **Sustainable Development**: Understanding the contributions of various economic sectors to a nation’s financial well-being is essential for achieving sustainable development.
* **Mitigating Economic Disruptions**: Considering current global uncertainties, marked by political changes, technological advancements, and the lasting impacts of the COVID-19 pandemic, this analysis becomes even more critical for mitigating economic disruptions and promoting economic resilience.

**Research Questions and Objectives:**

1. **Identify the Key Economic Drivers:**

* Objective: To determine which economic sectors or indicators are the most significant drivers of economic growth and stability in different countries
* Method: Implement clustering techniques to classify countries based on their economic attributes and analyze the shared characteristics within each group.

1. **Predict Economic Resilience**:

* Objective: To predict a country’s economic resilience based on historical economic indicators.
* Method: Utilize predictive modeling methods such as regression or time-series analysis to estimate future economic indicators and evaluate financial robustness.

1. **Evaluating Trade’s Role in Economic Stability**:

* Objective: To examine the relationship between a country’s trade balance (exports versus imports) and its economic stability and growth.
* Method: Apply classification algorithms to categorize countries based on their trade balance and study the ensuing impact on their economy.

1. **Analyzing Sector Contributions to GDP**:

* Objective: To investigate how different economic sectors contribute to the Gross Domestic Product (GDP), pinpointing those that are vital for economic stability.
* Method: Conduct correlation analysis and predictive modeling to understand the relationships between various sectors and GDP, and their roles in fostering economic stability.

# Data gathering and data analysis:

## Feature Selection:

We will now proceed with the literature review focusing on the key economic indicators identified from the dataset.

**Literature Review Focus:**

**1. Gross Domestic Product (GDP)**: Learn about the importance that GDP plays in forecasting economic growth as well as how to utilise it to measure economic performance and resilience.

**2. Gross National Income (GNI)**: Explore how GNI differs from GDP and its importance in economic analysis.

**3. Sectoral Outputs**: Investigate how different economic sectors contribute to overall economic stability and growth.

**4. Trade Balances**: Analyze the impact of trade (exports vs. imports) on economic health and resilience.

**Literature Review:**

**Reference 1: A comparison of monthly global indicators for forecasting growth.**

**1. GDP Growth and Economic Resilience:**

The study emphasizes how important GDP is as a thoroughly monitored metric, crucial for keeping tabs on macroeconomic trends and understanding the business cycle. Because of this, GDP is viewed as a key sign of economic resilience. Various models, like MIDAS and U-MIDAS, are used to nowcast and forecast quarterly real GDP growth worldwide.

("One of the most closely followed and comprehensive measures for monitoring macroeconomic developments is real GDP since it constitutes the primary indicator of the business cycle.”) [[1]](#_Bibliography:).

**2. Sectoral Output and GDP Contributions:**

The paper delves into sectoral output, shedding light on crucial economic sectors such as manufacturing, mining, and the electric and gas industries. This detailed analysis is instrumental for evaluating the role of sectoral contributions to GDP, aligning with the fourth research objective.

("The first economic indicator measures the level of real output in the manufacturing sector, mining, and the electric and gas industries worldwide”) [[1]](#_Bibliography:).

**3. Trade Balance and Economic Health:**

The relationship between trade balance, global economic activity, and the shipping industry is intricately explored, demonstrating how an increase in global economic activity can lead to a heightened demand for shipping services. This, in turn, raises shipping costs due to the short-term fixed supply of ships, providing valuable insights into how trade balance impacts economic stability and growth.

("The reasoning underlying this measure is that raw industrial materials need to be shipped before they can be used in production. An upswing in the global economy will lead to an increase in the demand for industrial commodities and thus shipping services, which raises the cost of shipping given that the supply of ships is fixed in the short run”) [[1]](#_Bibliography:).

**Reference 2: Components of GDP.**

**1. Sectoral Output and GDP Contributions:**

The study provides a detailed breakdown of the GDP by various sectors. It highlights the dominance of the service sector, accounting for approximately 82% of the output in 2021, and notes the declines in manufacturing and construction sectors in August 2023. This shows the importance of Sectoral Output towards the stability of the economy.

("Services are the largest part of the economy – making up about 82% of output in 2021. Manufacturing output fell by 0.8% in August 2023. Manufacturing is part of the wider production sector; production sector output fell by 0.7% in August 2023. Construction sector output decreased by 0.5% in August 2023.") [[2]](#_Bibliography:).

**Reference 3: Effects of COVID-19 on trade flows.**

**1. Trade Balance and Economic Health**:

The COVID-19 pandemic has underscored the critical relationship between trade balance and economic health, revealing significant decreases in trade flows, particularly in exports.

("We find that the COVID-19 pandemic has significantly decreased trade flows, with a larger effect on exports than imports") [[3]](#_Bibliography:).

("The disruption of global supply chains due to the pandemic has significantly affected the trade balance of countries") [[3]](#_Bibliography:). These insights align with our research objectives to assess the impact of trade on economic health and predict economic resilience, providing a foundation for developing strategies to enhance economic stability and growth in turbulent times.

**Reference 4: GDP and modified GNI.**

**1. GNI (Gross National Income):**

GNI serves as a more stable and reliable indicator, especially when modified to exclude disproportionate impacts of globalisation activities, as seen with the introduction of modified GNI (GNI\*) in Ireland.

("One of the key recommendations of the Group was for the CSO to develop a new indicator of the size of the economy that excludes globalisation activities that have a disproportionate impact on the Irish results. In July 2017, the CSO published an alternative measure of the size of the economy, so-called ‘modified Gross National Income’ (sometimes called GNI\*).") [[4]](#_Bibliography:).

Modified GNI provides a better approximation of the size of the economy and is deemed crucial for fiscal purposes and ratio analysis, aligning with our objective to accurately assess economic stability and predict resilience.

**2. Sectoral Output:**

The sectoral output, particularly from the aircraft leasing and IP sectors, plays a significant role in the economic indicators of countries with substantial multinational activities.

("Ireland is an important hub for multinational firms engaged in aircraft leasing, with around 50 percent of the world’s leased commercial aircraft managed here...this sector has a significant impact on the level of GNP / GNI.") [[4]](#_Bibliography:).

Understanding the contributions of different sectors to the overall economy is essential for our analysis, helping to identify key economic drivers and assess the role of sectoral contributions to GDP.

**Reference 5: Impact of selected sectors contribution to overall GDP of the Indian economy with reference to agricultural, industrial and service sector.**

**1. Sectoral Output:**

(“The agricultural sector provides food, raw materials, and employment to the people.") [[5]](#_Bibliography:).

("The Service Sector economic activities contribute to the growth of the agricultural and industrial sectors, which include trade, transport, communication, banking, education, health, tourism, services, insurance, etc.") [[5]](#_Bibliography:).

The above citations from the research paper prove that the sectoral output is a significant driver of economic stability and resilience, which is significant for identifying safe investment avenues and allocating resources effectively in times of economic risk.

Through the above-mentioned references, we can conclude that we can proceed with the indicators that have been chosen.

## Data Gathering:

The dataset was procured from Kaggle and originally compiled and maintained by the CDC’s Division of Population Health.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **Data Type** | **Data Level** | **Data Time** | **Data Structure** |
| GDP | Quantitative, Continuous | Ratio | Time-series | Numerical values |
| GNI | Quantitative, Continuous | Ratio | Time-series | Numerical values |
| Sectoral Output | Quantitative, Continuous | Ratio | Time-series | Numerical values |
| Trade Balance (Exports – Imports) | Quantitative, Continuous (Calculated) | Ratio | Time-series | Numerical values |

Figure 1: Overview of Selected Economic Indicators and Their Characteristics

## Data Analysis:

The data analysis framework is crafted to address the research questions and objectives.

**1. Data Pre-Processing**

Ensuring data integrity and readiness for analysis:

* 1. **Data Cleaning**: Address missing values and outliers in our key indicators. For instance, ensuring that all entries for GDP, GNI, Sectoral Output, and Trade Balance are present and accurately recorded.
  2. **Data Reduction**: Identify and retain the most relevant years or periods for analysis, possibly focusing on times of economic significance.
  3. **Proximity Analysis**: Examine the relationships and similarities between different countries based on their economic indicators to inform the clustering process.

**2. Data Processing**

Applying analytical methods to derive insights:

**2.1 Clustering (K-Means Clustering)**:

* Objective: Group countries based on their economic stability and growth potential.
* Indicators used: GDP, GNI, Sectoral Output
* Application: Identify patterns and commonalities among different groups of countries, shedding light on the key drivers of economic stability.
  1. **Classification**:
     1. **K-Nearest Neighbors:**
* Objective: Categorize countries based on their trade balance and its impact on economic health.
* Indicators used: Trade Balance, GDP
* Application: Analyze how trade surpluses or deficits correlate with economic stability and growth.
  + 1. **Decision Tree:**
* Objective: Understand the decision rules that lead to different economic states based on various indicators.
* Indicators: GDP, GNI, Sectoral Output, Trade Balance
* Application: Build a decision tree model to uncover the rules and conditions leading to economic resilience or vulnerability.
  1. **Predictive Modeling**:
     1. **Linear Regression:**
* Objective 1: Predict future economic resilience based on historical data.
* Objective 2: Understand how different economic sectors contribute to GDP.
* Indicators: GDP, GNI, Sectoral Output, Trade Balance
* Application: Forecast future values of GDP and GNI and model the relationship between GDP and Sectoral Output.
  + 1. **Time Series Analysis (ARIMA)*:***
* Objective: Forecast future values of economic indicators based on their historical patterns.
* Indicators: GDP, GNI, Trade Balance
* Application: Implement ARIMA models to capture and forecast the time-dependent patterns in the economic indicators.

# Conclusion:

In wrapping up this proposal, our main aim has been to dig deep into various economic indicators to pinpoint key contributors to economic growth and stability. By integrating clustering techniques, classification methods, and predictive analytics into our approach, we aspire to uncover vital patterns and relationships that can provide strategic decision-making for governments, investment firms, and global agencies.

Speaking about the outcomes, the K-Means clustering is expected to group nations based on their economic traits, highlighting the shared characteristics within each group. On the other hand, K-Nearest Neighbors and Decision Trees for classification will allow us to explore how trade balances influence a nation’s economic health. Furthermore, Linear Regression Time Series Analysis (ARIMA) for predictive modeling will enable us to project future economic indicators, facilitating the evaluation of a country's economic resilience.

However, it's important to note that this study has its constraints. The dataset, despite being extensive, might not fully reflect the complexities of global economies. Our dependence on secondary data and the potential encounter with outdated or incomplete information could impact the accuracy of our findings. Moreover, even though our chosen data analysis methods are effective, they might not be perfectly adaptable to every aspect of the economic event we are studying. Recognizing these constraints is important as we analyze our results and draw conclusions from our analysis.

# Bibliography:

1. Baumeister, C. and Guérin, P., 2021. A comparison of monthly global indicators for forecasting growth. *International Journal of Forecasting*, 37(4), pp.1276–1295. Available at: <https://doi.org/10.1016/j.ijforecast.2021.02.008>.
2. Harari, D., 2023. Components of GDP. *Economic Indicators*, (02787), p.1. Available at: <https://researchbriefings.files.parliament.uk/documents/SN02787/SN02787.pdf>.
3. Barbero, J., de Lucio, J. J., and Rodríguez-Crespo, E., 2021. Effects of COVID-19 on trade flows: Measuring their impact through government policy responses. *PLOS ONE*, 16(10), e0258356. Available at: <https://doi.org/10.1371/journal.pone.0258356>.
4. Economic Division, 2018. GDP and ‘modified GNI’ – explanatory note. *Parliamentary Budget Office*. Available at: <https://assets.gov.ie/4910/181218123252-71a2c297f26b419fa3696d7349e3e788.pdf>.
5. Jaiswal, M., 2022. Impact of Selected Sectors Contribution to Overall GDP of the Indian Economy with Reference to Agriculture, Industrial and Service Sector. In: *Sustainability of Business during COVID-19*. Shri Chinai College of Commerce and Economics: National Conference, p.80. Available at: <https://www.ijcrt.org/papers/IJCRTM020011.pdf>.