

# " Mapping Economic Similarities: Year Classification Based on Economic Inflation Indicators - ML "

## 1. Introduction

The main objective of this project was to identify economic similarities and inflation patterns between the years 1976 and 2023 in Bolivia. For this purpose, data provided by the World Bank was utilized, and machine learning techniques were applied to classify years that shared common economic and social patterns. The importance of this analysis lies in understanding how recurring economic events and implemented policies have impacted the Bolivian economy over time. Techniques such as K-means, Agglomerative Clustering, and Principal Component Analysis (PCA) were employed to create groups of years that exhibited similar behaviors, aiming to generate valuable insights for future economic policies.

## 2. Objectives

### General Objective:

Classify the years from 1976 to 2023 in Bolivia based on economic inflation indicators using machine learning algorithms.

### Specific Objectives:

- Identify common patterns in key inflation indicators such as the consumer price index, total debt service, trade balance, and interest rates.
- Analyze the interaction between economic policies and social events.
- Generate clusters grouping years with similar economic characteristics.

## 3. Methodology

### 3.1 Data

The dataset used was sourced from official World Bank data (<https://databank.worldbank.org/source/world-development-indicators>) and included the following indicators for Bolivia from 1976 to 2023:

- Consumer price index (2010 = 100)
- Inflation (annual %)
- GDP deflator (annual %)
- Lending interest rate (%)
- Real interest rate (%)
- Total debt service (% of exports of goods, services, and primary income)
- Total debt service (% of GNI)
- Current account balance (% of GDP and BoP, current US\$)
- Trade (% of GDP)

- Population, total

### 3.2 Algorithm

Both K-means and Agglomerative Clustering algorithms were applied for clustering analysis, with the latter being selected for its superior performance. This is a hierarchical method where each data point starts as an individual cluster, and in each iteration, the most similar clusters are merged until a predefined number of clusters is reached, or the similarity can no longer be reduced.

### 3.3 Data Preprocessing

Though the amount of missing data was minimal, missing values were completed using the arithmetic mean of the respective attribute or column to ensure data quality. Data normalization was also performed to ensure that all indicators carried equal weight in the analysis.

### 3.4 Implementation

The project was implemented in Python using the following libraries:

- **Pandas** for data loading and manipulation.
- **Scikit-learn** for executing the clustering algorithm.
- **Matplotlib** for visualizing the results.

### 3.5 Model Validation

The clustering model was validated using the following metrics:

- **Silhouette Coefficient:** Measures how similar a point is within its own cluster compared to other clusters.
- **Intra-cluster distance:** Measures the internal coherence of the formed clusters.
- **Calinski-Harabasz Index:** Measures the ratio between the dispersion among clusters and the dispersion within clusters.

## 4. Results

The results obtained from the clustering analysis allowed the grouping of years into different clusters that reflected common macroeconomic characteristics.

### 4.1 Year Clusterization

Five main clusters of years were identified, each showing significant similarities in terms of economic indicators:

- **Cluster "2" (1976-1986, except 1985, and including 2001 and 2002):** These years were characterized by military dictatorships and high inflation, with austerity policies that heavily impacted the population.
- **Cluster "3" (1985):** A year marked by hyperinflation, which was the result of both previous trends and populist policies of the ruling party, Unidad Democrática Popular.
- **Cluster "1" (1987-1999):** A period of neoliberal reforms, privatizations, and post-hyperinflation economic stabilization.
- **Cluster "0" (2000-2014, including 2020-2022):** A period of economic growth driven by resource nationalization and favorable commodity prices. The years 2020 to 2022 reflect an economic rebound after major financial stimulus packages.

- **Cluster "4" (2015-2023):** A period of economic slowdown due to the decline in commodity prices, a trade imbalance between imports and exports, rising internal debt, and the effects of the COVID-19 pandemic.

#### **4.2 Relation to Political and Social Events**

The clustering of years also reflects the impacts of political events, such as the return to democracy in 1982, economic liberalization policies in 1985, the "Gas War" in 2003, and the 2008 global economic crisis, which influenced Bolivia's economy and contributed to significant policy changes.

#### **5. Conclusions**

Through the application of machine learning and clustering techniques, it was possible to identify patterns in Bolivia's economic indicators over the last nearly 50 years. The results of this analysis provide a useful tool for understanding how different periods in Bolivia's history share similar economic characteristics, allowing policymakers to anticipate potential crises or replicate past successes in economic management.

Key conclusions include:

- The importance of political stability for sustained economic growth.
- The identified clusters confirm that periods of structural reforms (such as resource nationalization or privatization) marked significant economic shifts.
- This type of analysis could be expanded to analyze how Bolivia responds to future economic crises or to predict the impact of new fiscal and monetary policies.