Final Project Proposal: Machine Learning

I. Project Title

Clustering Regional Inflation Trends Using K-means for Economic Decision-Making in Google Colab

---

II. Team Information

Team Name (if applicable):

Team Members:

- Jolander C. Guiral
- James Byran V. Balambao
- Kent Remar L. Cunado

III. Background and Motivation

Problem Statement:

- How can unsupervised machine learning algorithms, like K-means clustering, help classify regions based on inflation trends and other economic indicators to inform strategic decision-making?
- How can advanced data analysis techniques, like clustering, uncover hidden patterns in regional inflation data to improve economic forecasting?
- What insights can clustered inflation data provide to help policymakers prioritize regions for economic aid and stabilize cost-of-living disparities?

Motivation:

- ***Inflation*** has a direct impact on the cost of living, consumer purchasing power, and overall economic stability. By identifying regions with similar inflation trends, policymakers can develop targeted interventions to address cost-of-living disparities and mitigate the negative impacts of high inflation in specific regions. This approach helps prevent widespread economic inequality and ensures that aid reaches the areas that need it the most.
- Understanding inflation trends through clustering can aid businesses and investors in making informed decisions about where to invest or expand. Regions with stable

inflation rates present opportunities for sustainable growth, while high-inflation regions may require special strategies to handle increased costs. Investors can use these insights to forecast market behaviors and adjust their portfolios accordingly.

VI. Objectives

- To explore and preprocess datasets effectively based on inflation trends and related economic indicators;
- To design and evaluate machine learning models using Python that highlight inflation trends and disparities among regions;
- To deploy and demonstrate the solution using Google Colab for effective communication of findings by stabilizing cost-living variations.

V. Scope

Target Users: Who will benefit from the solution?

- Policymakers and Government Agencies
- Economists and Researchers
- Business and Investors
- Non-Governmental Organizations
- Educators and Students

Features:

- Implementation of machine learning models, specifically the k-means clustering.
- Hyperparameter tuning and performance evaluation.
- Exploratory Data Analysis (EDA) and visualization
- Clustering Interpretation and Insights
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
- Use of Google Colab for easy collaboration and visualization.

VI. Dataset

Dataset Description: ***Kaggle*** is one of the best platforms for acquiring high-quality datasets, as they are curated by professionals and enthusiasts. It offers a wide range of topics, from finance and healthcare to social sciences and economics. The variety of datasets is highly relevant to today's industries, providing valuable solutions for handling various challenges. Kaggle also supports API access, allowing datasets to be directly

imported into Google Colab without any difficulties. Therefore, it is the best choice for extensive analysis and visualizing data structures to tackle meaningful problems in modern society.

Attributes/Features: Briefly describe the key features in the dataset.

- **Regional Member**: Describes the economic region or group to which a country belongs. This feature helps categorize inflation data based on regional economic blocks.
- **Year:** Represents the year in which the inflation data was recorded. This feature is critical for analyzing inflation trends over time and conducting temporal comparisons.
- **Inflation:** The inflation rate for a particular country in a specific year, typically expressed as a percentage. This is the primary target variable for clustering and analysis, as it directly impacts economic conditions.
- **Unit of Measurement:** Specifies the measurement unit used for inflation (e.g., percentage). This feature ensures that inflation data is appropriately interpreted and compared across different entries.
- **Subregion:** A more granular categorization of regions, indicating a specific subset of a broader region. This feature is important for regional economic analysis and understanding geographic disparities in inflation.
- **Country Code:** A unique identifier for each country in the dataset, typically in the form of a two-letter country. This helps link inflation data to specific countries and enables cross-country comparisons.

VII. Methodology

Tools and Libraries:

- **Python**: The core programming language for the project, enabling data processing, machine learning, and visualization tasks.
- **Matplotlib**: A powerful library for creating static, interactive, and animated visualizations in Python. It will help you create plots, graphs, and charts to visualize inflation trends, clustering results, and other insights from the data.

Model Selection:

- **K-means Clustering**: It is an unsupervised learning algorithm that will help classify regions based on inflation trends and other economic indicators. By grouping data into clusters, K-means will reveal patterns and similarities in inflation behaviors across different regions. This can assist in understanding economic disparities and aid in decision-making processes.

Steps:

1. Data preprocessing
2. Feature Scaling, Outlier Detection, and Categorical Data.
3. Use Line charts, Bar Graphs, or Heatmaps
4. Cluster Profiling
5. Exploratory Data Analysis (EDA) for insights and feature engineering.
6. Model training and evaluation using metrics like accuracy or precision.