**Research Proposal – Capstone Project Fall'23 (Team Argocd)**

**Title: *"Electrification Dynamics: Assessing Growth Patterns and Economic Impacts in Low-Income Countries, Benchmarking against Developed Nations"***

**Aim:** Develop a machine learning model to estimate the future economic inequality between developing (African) and developed countries based on their level of electrification.

**Synopsis:** Global Household Electrification Analysis: Classification and Prediction of Growth Rates in Low-Income Countries with a Focus on Electrification's Impact on Economic Indicators, Studying Developed Countries as Benchmarks.

**Introduction**

The global landscape of household electrification plays a pivotal role in socio-economic development. This project aims to conduct a comprehensive analysis of global household electrification using datasets obtained from the Harvard Dataverse and others from World Bank and IMF. The objective is to explore intricate relationships between household electrification rates and key economic indicators, such as Population, GDP, Socio-Economic Factors, Sustainable Development, Renewable Energy Development Opportunities, and Unemployment Rate. Understanding the dynamics between electrification and economic development is crucial for formulating effective policies and interventions. By leveraging machine learning techniques, we seek to classify countries based on their electrification status and predict how economic gap between countries (an African vs a developed country) will grow based on level of electrification.

**Research Gap**

In reviewing the previous team's projects, presentation videos, and judges' comments, we identified several key areas of focus. The teams delved into mapping areas with poor electrification, exploring cost-effective strategies for improvement, optimizing resource utilization, and constructing models for universal electrification. Moving forward, our research aims to build upon the foundation laid by previous teams; aiming to bridge the economic gap between developing (African) and developed countries. The economic gap between developing and developed countries based on level of electrification can be lessened if known beforehand (predict it) and the developing countries will apply the suggested recommendation(s). Our approach involves assessing potential economic prosperity indicators such as GDP, Unemployment Rate, Population, etc. We plan to compare these indicators with those of fully electrified households in developed countries. Through predictive analysis of electrification growth and the construction of corresponding models, our goal is to gain valuable insights into the potential economic prosperity of developing countries (African) associated with enhanced electrification benchmarking against developed countries.

**Link to Previous Research**

Our project builds on the work of previous teams by focusing on the economic implications of electrification. While previous teams ([Team Django](https://github.com/Dev-Sooraj-Git/ESP-HDSC-CAPSTONE-PROJECT.git), and [Team GCP](https://github.com/OBIGOD3931/Team-GCP-Hamoye-HDSC-Spring-23')) have mapped areas with poor electrification and explored strategies for improvement, our project takes a step further by examining the economic impact of these improvements.

We aim to address the criticisms of previous research:

* The judges queries why Team Django didn't develop any Regression model since the problem to be solved is a classification problem.
* Team GCP challenges in finding additional datasets.

We are going to address the above criticisms by:

* Accessing all the datasets needed for our project, which is already available.
* Providing a more comprehensive analysis of the relationship between electrification and economic development and by using ARIMA and LSTM model to predict future economic gaps between developing and developed countries based on level of electrification.

Our research will enhance the methodologies used in previous projects by incorporating machine learning techniques to predict future economic disparities based on electrification levels.

**Objectives**

1. Classify countries based on their electrification status using supervised learning algorithms.
2. Explore correlations between electrification rates and key economic indicators such as Population, GDP, Unemployment rate, etc.
3. Predict how the economic gap between countries (an African vs a developed country) will grow based on level of electrification.

**Methodology**

**A. Data Collection**

* Utilize the provided dataset on Global Household Electrification[.](https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/BZBKJP/UTKTNH&version=2.1&toolType=QUERY)
  + [Global Electrification Database](https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/BZBKJP/UTKTNH&version=2.1&toolType=QUERY)
* Integrate additional datasets on Population, GDP, CPI, Unemployment rate, and other relevant economic indicators.
  + [Unemployment dataset](https://www.imf.org/external/datamapper/LUR@WEO/OEMDC/ADVEC/WEOWORLD)
  + [GDP by Country](https://www.kaggle.com/datasets/tunguz/country-regional-and-world-gdp)
  + [CPI dataset](https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG)
  + [Countries’ population and Income Group Dataset](https://data.worldbank.org/indicator/SP.POP.TOTL)
* Ensure a common identifier for merging datasets and maintaining data integrity.

**B. Exploratory Data Analysis (EDA) and Visualisation**

* Conduct EDA to understand the distribution and patterns within the data. Identify outliers, trends, and potential opportunities for feature engineering.
* Create visualizations (Maps, Charts, Graphs) to effectively communicate findings.

**C. Machine Learning Models**

* Employ supervised learning algorithms - Decision Trees and/or Random Forests to classify countries based on their electrification status.
* Investigate the relationship between electrification rates and economic indicators.
* Use statistical methods to quantify the strength and direction of correlation.
* Build Time Series Forecasting Models using ARIMA and LSTM machine learning algorithms to predict how the economic gap between countries (an African vs a developed country) will grow based on level of electrification.

**Deliverables**

* Classification models for electrification status.
* Visualizations to convey complex relationships and patterns.
* Correlation analyses between electrification rates and economic indicators.
* Predictive models for economic gaps between countries (an African vs a developed country) will grow based on the level of electrification.

**Recommendations**

These recommendations aim to bridge the gap between electrification efforts and economic prosperity, fostering sustainable development in low-income countries while drawing valuable insights from the benchmarking analysis against developed nations. Based on this we propose the following recommendations:

* **Develop targeted electrification programs for low-income countries**: The research could identify the specific factors that are hindering electrification in low-income countries, and these findings could be used to develop targeted programs that address these challenges.
* **Resource Optimization for Economic Benefits:** Develop strategies to optimize resources by aligning electrification projects with key economic indicators. This involves aligning electrification efforts with population growth, GDP targets, and sustainable development goals to maximize economic benefits.
* **Renewable Energy Integration:** Encourage the integration of renewable energy sources in electrification projects. The correlation analysis provides an opportunity to identify regions where renewable energy viability aligns with electrification growth. Recommendations can be made to prioritize these areas for sustainable and economic development.

The research findings could also be used to advocate for increased funding for electrification initiatives. By demonstrating the economic benefits of electrification, the research could help to convince governments and investors to invest more money in this critical area of development.

**Validation and Sensitivity Analysis**

Validate models and conduct sensitivity analysis to ensure robustness.

**References**

* datacatalog.worldbank.org. (n.d.). *Global Electrification Platform (GEP) Application | Data Catalog*. [online] Available at: https://datacatalog.worldbank.org/search/dataset/0037768/Global-Electrification-Platform--GEP--Application [Accessed 20 Nov. 2023].
* ‌ Nwachukwu, O. (2023). *Team-GCP-Hamoye-HDSC-Spring-23-Capstone-Project*. [online] GitHub. Available at: https://github.com/OBIGOD3931/Team-GCP-Hamoye-HDSC-Spring-23-Capstone-Project [Accessed 20 Nov. 2023].
* ToluAnnie (2023). *ToluAnnie/global-household-electrification-project*. [online] GitHub. Available at: https://github.com/ToluAnnie/global-household-electrification-project [Accessed 22 Nov. 2023].
* SHET, S. (2023). *Dev-Sooraj-Git/ESP-HDSC-CAPSTONE-PROJECT*. [online] GitHub. Available at: https://github.com/Dev-Sooraj-Git/ESP-HDSC-CAPSTONE-PROJECT [Accessed 21 Nov. 2023].
* ‌www.youtube.com. (n.d.). *HDSC Winter ’23 Grandfinale Presentations*. [online] Available at: https://www.youtube.com/watch?v=jNhCHfHwFd8 [Accessed 21 Nov. 2023].