**A blue and white flag with a logo

Description automatically generated**

**Frontier Tech** Leaders **Global Cohort Machine Learning Bootcamp #2**

**Title:**

Reducing Urban Poverty through

Economic Data Analysis

**Group9**

BY

Osamah SHARAF ALDEEN

Hadeel TARQI

**Capstone Project Concept Note and Implementation Plan**

**Project Title:** Reducing Urban Poverty through Economic Data Analysis

**Team Members:**

1. Osamah SHARAF ALDEEN
2. Hadeel TARQI

**Concept Note**

**1. Project Overview**

This capstone project aims to predict socio-economic indicators for various countries, contributing to the Sustainable Development Goals (SDGs), particularly SDG 1 (No Poverty) and SDG 8 (Decent Work and Economic Growth). By accurately predicting economic indicators, policymakers can make informed decisions to foster economic growth and reduce poverty. The project's potential impact includes better resource allocation and improved socio-economic planning at the national and international levels.

**2. Objectives**

* Develop a machine learning model to predict socio-economic indicators for countries.
* Provide a user-friendly interface for stakeholders to input data and receive predictions.
* Contribute to data-driven policymaking by offering accurate and timely economic predictions.

**3. Background**

The project addresses the need for reliable economic forecasting tools to support SDG goals. Existing solutions often rely on traditional statistical methods, which may not capture complex relationships in the data. A machine learning approach can identify patterns and provide more accurate predictions. While initiatives like the World Bank's economic indicators provide valuable data, integrating machine learning enhances prediction accuracy and utility.

**4. Methodology**

The project will use machine learning techniques such as RandomForestClassifier for predictive modeling. Key steps include data preprocessing, feature engineering, model training, and evaluation. The use of GridSearchCV for hyperparameter tuning and cross-validation ensures robust model performance.

**5. Architecture Design Diagram**

* **Data Collection:** Gather socio-economic data from various sources.
* **Data Preprocessing:** Clean and transform the data for modeling.
* **Feature Engineering:** Create and select relevant features for the model.
* **Model Training:** Train machine learning models using Scikit-Learn.
* **Model Evaluation:** Evaluate model performance using cross-validation.
* **Deployment:** Deploy the model using Gradio for a user-friendly interface.

**6. Data Sources**

The primary dataset, "PovStatsCountry.csv," is sourced from the World Bank, containing various socio-economic indicators for different countries. The data is in CSV format with 184 rows and 30 columns. Preprocessing steps include handling missing values, encoding categorical variables, and normalizing numerical features.

**7. Literature Review**

In the forthcoming sections of this literature review, I will expand on several critical areas. First, I will delve deeper into the methodologies employed in the referenced studies, providing a comprehensive analysis of their approaches and how they can be adapted for our research. Next, I will synthesize findings from the three thematic areas, identifying common patterns and unique insights that contribute to understanding urban poverty. Following this, I will conduct a comparative analysis to highlight the interplay between financial inclusion, economic growth, and education in mitigating urban poverty. Additionally, I will discuss potential policy implications derived from the literature, offering practical recommendations for policymakers. Finally, I will outline the limitations of existing studies and propose directions for future research to fill identified gaps and further the field.

**Implementation Plan**

**1. Technology Stack**

* **Programming Languages:** Python
* **Libraries:** Scikit-Learn, Gradio, MLflow, Pandas, NumPy, Seaborn, Matplotlib
* **Frameworks:** Jupyter Notebook, Flask (for potential web deployment)
* **Tools:** GitHub, JupyterLab, Joblib for model saving and loading

**2. Timeline**

* **Week 1-2:** Data collection and preprocessing
* **Week 3-4:** Model development and initial training
* **Week 5-6:** Model evaluation and hyperparameter tuning
* **Week 7:** Deployment and user interface development
* **Week 8:** Final testing and project submission

**Gantt Chart:**

**Task Distribution Matrix:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Task** | **Team Member 1** | **Team Member 2** |  |  | | --- | --- | --- | --- | --- | | Data Collection | X |  |  |  | | Data Preprocessing |  | X |  |  | | Model Development | X |  |  |  | | Model Training |  | X |  |  | | Hyperparameter Tuning | X |  |  |  | | Deployment |  | X |  |  | | User Interface Development | X |  |  |  | | Final Testing | X | X |  |  | |

**3. Milestones**

* **Milestone 1:** Completion of data preprocessing
* **Milestone 2:** Development of initial machine learning model
* **Milestone 3:** Successful hyperparameter tuning
* **Milestone 4:** Deployment of the model with a functional user interface
* **Milestone 5:** Final evaluation and testing of the model

**4. Challenges and Mitigations**

* **Data Quality:** Ensure thorough data cleaning and preprocessing to handle missing values and inconsistencies.
* **Model Performance:** Utilize cross-validation and hyperparameter tuning to optimize model performance.
* **Technical Constraints:** Leverage cloud resources if local computational power is insufficient.

**5. Ethical Considerations**

* **Data Privacy:** Ensure that the data used does not contain personally identifiable information (PII).
* **Bias:** Regularly evaluate the model for potential biases and take steps to mitigate them.
* **Impact:** Consider the socio-economic impact of the predictions and ensure they are used responsibly

**6. References**

* PEDREGOSA, Fabian, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 2011, 12: 2825-2830.‏

|  |
| --- |
| * ZAHARIA, Matei, et al. Accelerating the machine learning lifecycle with MLflow. *IEEE Data Eng. Bull.*, 2018, 41.4: 39-45.‏ * World Bank. "World Development Indicators." [Online]. Available:   . https://www.kaggle.com/datasets/theworldbank/poverty-and-equity-database |

* ABID, Abubakar, et al. Gradio: Hassle-free sharing and testing of ml models in the wild. *arXiv preprint arXiv:1906.02569*, 2019.‏