**Poverty Prediction and Analysis using AI**



Session: 2022 – 2026

**Submitted by:**

A.W Huzaifa Maalik Wattoo

2022-CS-145

M. Tayyab

2022-CS-135

**Submitted to:**

Mr. Samyan Qayyum Wahla

Department of Computer Science

**University of Engineering and Technology Lahore, Pakistan**

**Table of Contents**

1. Executive Summary
2. Introduction
   * Objective of the Study
   * Scope of the Project
3. Data Analysis
   * Overview of the Dataset
   * Data Preprocessing
4. AI Model and Methodology
   * Model Details
   * Feature Importance
5. Dashboard and Visualizations
   * Functionality
   * Implementation
6. Results and Discussion
   * Observations
   * Model Validation
7. Recommendations
8. References

**Executive Summary**

This project uses a machine learning-based Random Forest model to predict poverty levels in Pakistan. It employs socio-economic variables such as gender, age, unemployment rate, and region. Predictions are visualized on an interactive Streamlit dashboard to aid policy decisions aligned with Sustainable Development Goal 1 (SDG 1): No Poverty.

**Introduction**

**Objective of the Study**

The primary objectives of this study are:

* To identify key socio-economic factors influencing poverty in Pakistan.
* To develop a predictive model that helps policymakers address poverty challenges effectively.

**Scope of the Project**

This project focuses on:

* Leveraging socio-economic data to build a machine learning model.
* Providing real-time insights via a user-friendly interactive dashboard.
* Supporting data-driven decision-making for achieving SDG 1.

**Data Analysis**

**Dataset Overview**

The datasets used, **Poverty.xlsx** and **Unemployed.xlsx**, were merged to create an integrated socio-economic dataset. The datasets included:

* **Demographics:** Region, Sex, and Age.
* **Economic Indicators:** Unemployment rates and poverty levels.

**Preprocessing:**

1. Merging datasets on common attributes: Region, Sex, Age, and Year.
2. Removing null values and irrelevant columns.
3. Encoding categorical variables (Sex and Age) into numeric formats using label encoding.

**AI Model and Methodology**

**Model Details**

* **Algorithm:** Random Forest Regressor.
* **Hyperparameter Optimization:** Conducted using GridSearchCV to tune parameters like n\_estimators, max\_depth, and min\_samples\_split.
* **Performance Metrics:**
  + RMSE: 0.98
  + MAE: 0.12

**Feature Importance**

The model identified key features influencing poverty levels:

1. **Unemployment Rate:** Strongly correlated with poverty.
2. **Region:** Highlights economic disparities.
3. **Age Group:** Differentiates poverty risks among Youth, Adults, and Seniors.
4. **Gender:** Reflects societal and economic gender dynamics.

**Dashboard and Visualizations**

**Functionality**

1. **Real-Time Prediction:** The dashboard accepts user inputs (Sex, Age Group, Year, Unemployment Rate) and predicts poverty levels instantly.
2. **Impact Analysis:** Visualizes changes in poverty rates with varying unemployment rates.
3. **Feature Insights:** Displays feature importance contributing to poverty predictions.

**Implementation**

* **Code Structure:**
  + frontend.py: Implements the Streamlit dashboard.
  + model.py: Handles data preprocessing, model training, and evaluation.
* **Interactive Elements:** Allows users to input demographic data and explore predictions and their drivers.

**Results and Discussion**

### Observations

* **Unemployment:** High unemployment rates consistently correlate with elevated poverty levels across the analyzed data.
* **Age Groups:** Youth are disproportionately affected by poverty, indicating a need for targeted interventions.
* **Gender Dynamics:** Gender-based disparities suggest socio-cultural factors play a role in poverty outcomes.

**Model Validation**

* Predictions align with socio-economic realities, verified through performance metrics (low RMSE and MAE).

**Recommendations**

1. **Enhance Skill Development:** Focus on training youth in areas with high unemployment.
2. **Invest in Infrastructure:** Improve access to healthcare and education in underserved regions.
3. **Regular Updates:** Periodically update the dashboard with new data to ensure accuracy.

**References**

* **Data Source:** Derived from ILOSTAT database.
* **Datasets:** Poverty.xlsx, Unemployed.xlsx.
* **Project Files:**
  + frontend.py: Dashboard implementation.
  + model.py: Preprocessing and model training.
* **Machine Learning Model:** Random Forest Regressor, fine-tuned with GridSearchCV.