Project Report: Economic Data Analysis – Cost of Living Index Prediction

1. Project Overview

This project aims to analyze economic indicators and predict the Cost of Living Index for different countries using statistical and machine learning models. The data used consists of various indices such as Rent Index, Groceries Index, Restaurant Price Index, and Local Purchasing Power Index for the year 2022. The primary objective is to build an efficient regression model and deploy it using a Streamlit web application.

2. Dataset Description

Source: Cost_of_Living_Index_2022.csv

Number of Records: 139 (countries)

Features Used:

- Rent Index

- Groceries Index

- Restaurant Price Index

- Local Purchasing Power Index

Target Variable: Cost of Living Index

3. Data Preprocessing

- Loaded the dataset and removed missing/null entries.

- Converted relevant columns to numerical format.

- Verified feature correlations using a heatmap.

- Identified the most influential features for the prediction model.

4. Model Training

4.1 Linear Regression

Algorithm: Linear Regression from Scikit-learn

Features Used:

- Rent Index

- Groceries Index

- Restaurant Price Index

- Local Purchasing Power Index

Target: Cost of Living Index

4.2 Performance Metrics

- Mean Squared Error (MSE): Indicates average prediction error:12.07

- R-squared Score (R^2): Measures the proportion of variance explained by the model:0.97

4.3 Model Output

- The model was trained and saved using joblib for reuse in deployment.

- Feature coefficients and intercept were examined to understand feature impact.

5. Model Deployment

5.1 Streamlit Web Application

A user-friendly web interface was created using Streamlit to make predictions.

Features:

- User inputs via sliders for each index.

- Real-time prediction display.

- Option to view model coefficients and intercept.

5.2 Workflow

1. Load the pre-trained model using joblib.

2. Take user inputs.

3. Predict the Cost of Living Index.

4. Display results dynamically.

6. Files Included

- Cost_of_Living_Index_2022.csv – Dataset

- model.pkl – Trained regression model

- train_and_save_model.py – Script to train and save the model

-dashboard.py – Streamlit web application script

-economic_ project_report.pdf – This document

7. Future Improvements

- Add support for batch predictions using uploaded CSV files.

- Integrate more features (e.g., inflation rate, GDP per capita).

- Improve prediction accuracy using advanced models (e.g., Ridge, Random Forest).

- Deploy to cloud (e.g., Streamlit Cloud, Heroku).