**Linear Regression Model Documentation: Predicting Nigeria's Inflation Rate**

This documentation outlines the process of developing and evaluating a linear regression model to predict Nigeria's inflation rate using real-world economic data. The model leverages the relationship between various Consumer Price Index (CPI) components and key economic indicators to predict the inflation rate, and its performance is measured using the Mean Absolute Error (MAE) metric.

**Problem Statement**

* **Objective:**  
  To create a predictive model that can estimate Nigeria's inflation rate based on select features of various CPI components.
* **Goal:**  
  The model will take input values for selected economic features (like CPI components) and predict the inflation rate, providing a tool for forecasting and understanding economic trends.

**Data Description**

The dataset provides a comprehensive overview of Nigeria's monthly inflation rates from March 2003 to June 2024. It also includes key economic indicators such as:

* Crude oil prices
* Production levels
* Various Consumer Price Index (CPI) components

This dataset is ideal for time series analysis, forecasting, and economic modeling, capturing significant trends that are valuable for economic prediction.

**Model Training**

**Import Statements and Dataset**

To begin with, the necessary libraries and modules are imported, and the dataset is loaded into the model.

python

Copy

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import sklearn

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_absolute\_error

from sklearn.metrics import mean\_squared\_error, r2\_score

**Data Splitting**

The dataset contains both features (independent variables) and the target variable (inflation rate). Here, we use one feature, CPI\_Food, for model training.

* **X (features):** Consumer Price Index (CPI) for food
* **y (target):** Inflation rate

The data is split into training and test sets, with an 80%/20% ratio for training and testing, respectively:

python

Copy

X = data['CPI\_Food']

y = data['Inflation Rate']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

**Model Training**

The linear regression model is trained using the training set (X\_train and y\_train):

python

Copy

model = LinearRegression()

model.fit(X\_train.values.reshape(-1, 1), y\_train) # Reshaping X\_train for fitting the model

**Model Evaluation**

After training the model, predictions are made on both the training and test sets. The performance is evaluated using the **Mean Absolute Error (MAE)** metric, which measures the average magnitude of errors in a set of predictions. A lower MAE indicates better model accuracy.

* **Baseline Model MAE:** 3.45 (This is the performance of a simple guess, like predicting the average value of the target variable for all data points.)
* **Model (Train) MAE:** 2.42 (The MAE on the training data, showing how well the model fits the training set.)
* **Model (Test) MAE:** 2.33 (The MAE on the test data, showing how well the model generalizes to unseen data.)

The **test MAE of 2.33** demonstrates that the model’s predictions are closer to the actual inflation rates compared to the baseline model (MAE of 3.45), indicating an improvement in predictive accuracy.

python

Copy

# Make predictions on train and test sets

train\_predictions = model.predict(X\_train.values.reshape(-1, 1))

test\_predictions = model.predict(X\_test.values.reshape(-1, 1))

# Calculate MAE for training and testing sets

train\_mae = mean\_absolute\_error(y\_train, train\_predictions)

test\_mae = mean\_absolute\_error(y\_test, test\_predictions)

**Conclusion**

The model has demonstrated a significant improvement in predictive accuracy compared to a baseline model, achieving a **test MAE of 2.33**. This indicates that the linear regression model, trained on CPI data (in this case, CPI\_Food), is a useful tool for predicting Nigeria's inflation rate.

Further improvements could involve:

* Incorporating additional features (e.g., crude oil prices, production levels) into the model
* Evaluating more advanced models such as polynomial regression or machine learning models (e.g., Random Forest or XGBoost)

**Future Work**

Future work can focus on:

* Expanding the feature set to include more economic indicators.
* Evaluating model performance over longer time periods or using different time series techniques for more accurate forecasting.
* Experimenting with advanced models to improve prediction accuracy and handle non-linearity in the data.