U.S. Inflation Prediction Using Macroeconomic Indicators (2000–2023)

Serdar Hoşver
TED University, Graduate Program in Applied Data Science
Course: ADS 577 | Instructor: Asst. Prof. Dr. Hakan Emekci

---

1. Introduction & Research Objective

Inflation is a pivotal economic indicator that reflects the health of an economy and influences monetary policy decisions. Predicting inflation accurately is essential for central banks, fiscal authorities, and financial markets. This project aims to model and predict the **monthly U.S. inflation rate** using three key macroeconomic variables:

- **Federal Funds Rate** (interest rate policy)
- **Unemployment Rate** (labor market slack)
- **M2 Money Supply** (liquidity in the economy)

The study addresses the question:

"Can we predict next month's U.S. inflation rate based on these indicators?"

To answer this, a combination of statistical and machine learning models were employed on historical data sourced from the **Federal Reserve Economic Database (FRED)** covering the years 2000–2023.

---

2. Data and Methodology

2.1. Data Overview

The analysis used monthly time-series data for:

- **CPI (Consumer Price Index)**: Proxy for inflation
- **FEDFUNDS**: Federal Funds interest rate
- **UNRATE**: Unemployment rate
- **M2SL**: M2 money supply levels

All datasets were cleaned, merged on dates, and examined for missing values and stationarity. The merged dataset consisted of 277 observations spanning over two decades.

2.2. Exploratory Data Analysis

Visual and statistical analyses included:

- **Time-series plots** for CPI, interest rates, and M2.
- **Histograms** revealed skewed distributions for CPI and M2.
- **Correlation heatmap** indicated:
  - Positive correlation between M2 and inflation,
  - Negative correlation between interest rates and inflation.

Pairplots and 12-month rolling averages were also used to explore lag effects and trends across crises (2008, COVID-19).

---

3. Modeling and Results

Three regression models were built:

- **Linear Regression**
- **Random Forest Regressor**
- **Decision Tree Regressor**

The dataset was split into training and test sets (80/20). Performance was evaluated using MAE, MSE, and R^2 scores.

3.1. Model Performance

Model	MAE	MSE	R^2
Linear Regression	1.573	3.530	0.068
Random Forest	0.319	0.230	0.939
Decision Tree	0.496	0.488	0.871

- **Random Forest** yielded the best accuracy, indicating strong non-linear relationships in the data.
- **Linear Regression** underperformed due to inability to capture structural breaks or threshold effects.

3.2. Feature Importance

The Decision Tree model highlighted:

- **M2 Money Supply** as the most important predictor,
- Followed by **Federal Funds Rate**,
- And then **CPI**, suggesting persistence in inflationary dynamics.

3.3. Prediction Example (Latest Data)

Predicted inflation rate for the next month:

- Linear Regression: **7.77%**
- Random Forest: **10.09%**
- Decision Tree: **9.84%**

---

4. Discussion and Economic Implications

The results confirm well-established economic theories:

- Monetary expansion (M2 growth) and low interest rates can elevate inflation.
- Labor market indicators (like unemployment) have a delayed or weaker contemporaneous effect.

These insights can guide:

- **Federal Reserve** in adjusting interest rates proactively.
- **Fiscal policymakers** in gauging inflationary pressure before stimulus decisions.
- **Investors** in interpreting bond yields and currency value movements.

---

5. Conclusion

This project successfully demonstrated the utility of macroeconomic indicators in forecasting U.S. inflation. It also revealed the superiority of ensemble models like Random Forest in capturing non-linear relationships. The approach provides a practical framework for enhancing policy forecasting and contributes to the broader field of economic data science.