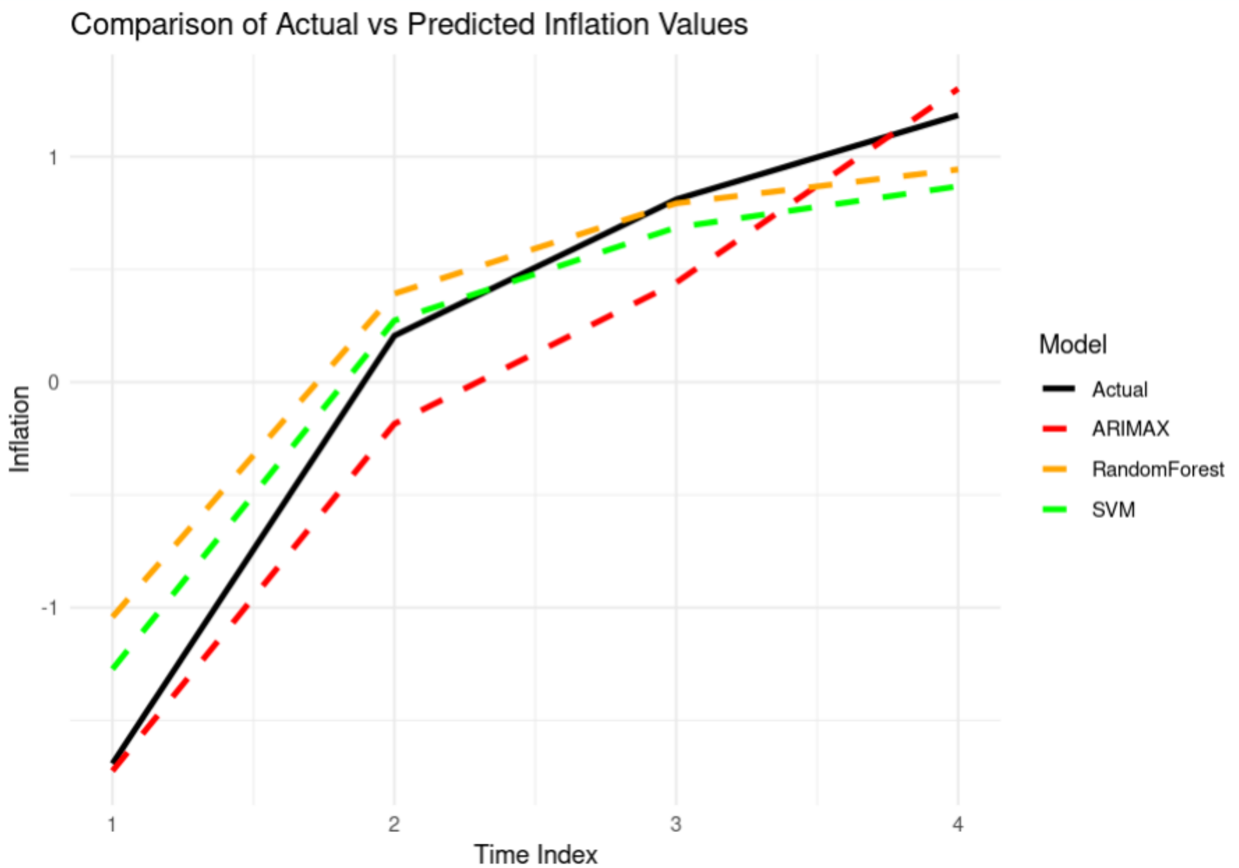
Advance Statistics

Prediction of USA' Inflation Using ML Models

ARIMA , SVM , Random Forest



Introduction

Inflation forecasting plays a pivotal role in economic planning, policy formulation, and decision-making. It refers to the process of predicting the rate at which the general level of prices for goods and services rises, eroding the purchasing power of money. Accurate inflation forecasts are crucial for governments, businesses, and financial institutions as they influence a wide range of economic decisions, including interest rates, wages, tax policies, and investment strategies.

Over the years, various models and methods have been developed to predict inflation, utilizing a range of data sources, from macroeconomic indicators to market expectations.

Context and Importance of Inflation Forecasting:

Inflation has significant economic implications. It directly impacts the purchasing power of consumers, affecting everything from household savings to the cost of living. For governments and central banks, managing inflation is a core economic objective. High inflation can lead to reduced consumer confidence, economic instability, and a decline in real incomes, while low inflation or deflation can signal weak demand and an economic slowdown. Hence, accurate forecasting of inflation becomes vital to ensure that policy interventions can be timely and effective.

Inflation forecasting is a complex task due to the many factors influencing price changes. These include demand and supply dynamics, labor market conditions, global economic factors, and fiscal policies. Moreover, inflation is inherently a lagging indicator, meaning it reflects past trends and conditions, making its forecasting even more challenging. Despite these challenges, inflation prediction is essential to maintain economic stability, as governments and central banks adjust their monetary policies based on forecasted inflation trends.

Historical Context:

The history of inflation forecasting dates back to the early 20th century when economists began recognizing the importance of anticipating price movements. Early approaches focused on simple statistical models that relied heavily on historical data. The simplest of these models, such as univariate time series models, relied solely on past inflation rates to predict future values. However, such models were limited in their accuracy, particularly during periods of significant economic change or external shocks.

The evolution of inflation forecasting models accelerated with the rise of macroeconomic theory and econometric techniques in the mid-20th century. Economists like Phillips and Keynesian scholars emphasized the relationship between inflation and unemployment, leading to the development of the Phillips Curve, which suggested that inflation could be managed through government fiscal and monetary policies. During this period, econometric models gained prominence as policymakers sought to understand the effects of monetary policies on inflation.

By the 1970s and 1980s, with the rise of globalization and increasing complexity in the global economy, traditional models became less effective. The 1973 oil crisis, for example, demonstrated that supply-side shocks could significantly alter inflationary trends, prompting economists to rethink the assumptions of many forecasting models. Newer approaches, such as structural macroeconomic models, began to incorporate more variables and detailed data sets to better capture the complexities of inflation dynamics.

In the 1990s and early 2000s, the increased availability of high-frequency economic data, coupled with advancements in computational power, paved the way for more sophisticated models. Central banks, including the Federal Reserve and the European Central Bank, began using a combination of various econometric and statistical models to forecast inflation more effectively. These methods allowed policymakers to refine their inflation-targeting strategies, making their interventions more precise and timely.

Science Behind Inflation Forecasting:

Inflation forecasting models can be broadly classified into two categories: **structural models** and **time-series models**.

1. **Structural Models:**
   Structural models, such as the New Keynesian model, aim to explain inflation based on underlying economic principles. These models incorporate relationships between various economic variables such as output, unemployment, interest rates, and wages. Structural models typically rely on economic theory to explain the channels through which different factors influence inflation. For example, the output gap—defined as the difference between actual and potential output—plays a significant role in predicting inflation. When the economy is operating above its potential (i.e., when there is high demand for goods and services), inflation tends to rise.

2. **Time-Series Models:**
   Time-series models, on the other hand, rely purely on historical data to predict future inflation. The most common time-series models include **Autoregressive Integrated Moving Average (ARIMA)** models, which capture the temporal dependencies in the inflation data. These models assume that future inflation can be predicted based on past inflation trends, with no explicit need to consider the underlying economic causes. However, time-series models can often be improved by including external variables, such as oil prices, exchange rates, and labor market indicators.

Over time, economists have combined elements of both approaches, leading to the development of **mixed-frequency models** and **forecast combination techniques**. These models incorporate multiple sources of data, such as expert judgment, macroeconomic variables, and combinations of various forecasting models, in an effort to improve accuracy. Additionally, machine learning techniques, such as **Support Vector Machines (SVMs)** and **Random Forest models**, have recently been applied to inflation forecasting, providing more advanced methods for capturing complex, nonlinear relationships in data.

Reason for Choosing the Topic:

The motivation for choosing the topic of inflation forecasting stems from its central importance in economic policymaking and financial planning. As the world continues to experience rapid economic transformations—driven by factors such as globalization, technological advancements, and climate change—the ability to predict inflation becomes even more critical. Traditional models, while valuable, often fail to account for the complexities introduced by global interconnectedness and unforeseen shocks. This has prompted a renewed interest in developing more robust, adaptive forecasting methods that can incorporate a wider range of variables and data sources.

The recent global economic disruptions, such as the COVID-19 pandemic and subsequent supply chain challenges, have further underscored the need for more accurate inflation predictions. The pandemic, for example, led to both supply-side and demand-side shocks that were difficult to predict using traditional forecasting models. Therefore, this research aims to contribute to the field of inflation forecasting by exploring advanced methodologies, including the use of machine learning techniques and forecast combinations, to enhance prediction accuracy. By incorporating a broader array of data sources and employing cutting-edge statistical methods, this study aims to improve our understanding of inflation dynamics and assist policymakers in making more informed decisions.

In conclusion, inflation forecasting remains a critical area of study due to its far-reaching implications for economic stability and policy effectiveness. As economies continue to face new challenges and uncertainties, the ability to predict inflation with greater accuracy will be essential for governments and financial institutions alike. This report, therefore, aims to investigate the current state of inflation forecasting, assess the effectiveness of various methodologies, and contribute new insights into how forecasting can be improved to meet the demands of the modern economic landscape.

Literature Review

Purpose of The Literature Review

The purpose of this literature review is to explore existing research and methodologies related to inflation forecasting, with a particular focus on the United States. In this review, we aim to analyze the strengths, weaknesses, and applications of various forecasting techniques, including traditional time series models such as ARIMA (AutoRegressive Integrated Moving Average) and machine learning models like Support Vector Machines (SVM) and Random Forest. By understanding the methodologies utilized in past studies, this review seeks to provide a foundation for selecting the most effective forecasting techniques and highlight gaps in current research that this project intends to address.

Contextualization

Inflation forecasting is a critical component of economic analysis, as it provides policymakers, businesses, and investors with valuable insights into future price trends. Accurate inflation predictions are essential for making informed decisions regarding monetary policy, interest rates, wage adjustments, and long-term investment planning. In developed countries like the USA, inflation forecasting plays a pivotal role in shaping the broader economic landscape. Given the complexity and volatility of inflationary pressures—often influenced by domestic and international factors—having reliable forecasting models can help mitigate economic risks and guide the formulation of effective policy responses. Moreover, inflation forecasts allow central banks, such as the Federal Reserve, to implement targeted measures to control inflation and ensure stable economic growth.

Scope of the Review

This literature review will cover the application of various inflation forecasting techniques, focusing on both traditional and modern methods. The primary models under review will include ARIMA, which is widely used for time series forecasting due to its ability to model trends and seasonality in inflation data. Additionally, the review will discuss machine learning techniques such as Support Vector Machines (SVM) and Random Forest, which have gained popularity in recent years for their capacity to handle large datasets and capture complex, non-linear relationships in economic data. The effectiveness of these models will be evaluated based on their ability to predict inflation in the context of the USA, with comparisons drawn from empirical studies that have applied these methods to similar macroeconomic forecasting tasks. Finally, the review will discuss the advantages and limitations of these methods, including considerations of model accuracy, computational complexity, and the potential for overfitting. This review aims to provide a comprehensive understanding of the state of inflation forecasting research, identifying key methodologies and highlighting areas for further exploration.

Theoretical Background on Inflation and Economic Forecasting

Inflation Definition

Inflation refers to the general increase in the price level of goods and services over time, resulting in a decrease in the purchasing power of a currency. It is typically measured as the percentage change in the Consumer Price Index (CPI) or other price indices over a specific period, such as monthly or annually. Inflation can be categorized into two types: **demand-pull inflation**, which occurs when demand for goods and services exceeds their supply, and **cost-**

push inflation, which arises when the cost of production increases (e.g., due to rising wages or raw material prices). A moderate level of inflation is often considered necessary for economic growth, but when inflation becomes too high or too volatile, it can lead to economic instability.

In developed economies like the USA, inflation is closely monitored and controlled due to its potential to impact multiple aspects of the economy. High inflation can erode consumer purchasing power, disrupt savings, and lead to uncertainty in investment decisions. On the other hand, deflation—where prices fall across the economy—can lead to reduced economic activity, higher unemployment, and potentially a recession. Therefore, central banks, such as the Federal Reserve in the USA, aim to maintain inflation at a target rate (usually around 2%) to ensure stable economic conditions, foster growth, and protect the value of the currency.

Inflation Forecasting

Inflation forecasting is a crucial tool for central banks, policymakers, and financial analysts. By accurately predicting future inflation rates, these stakeholders can take preemptive actions to stabilize the economy, adjust interest rates, and formulate appropriate fiscal and monetary policies. Effective inflation forecasting helps mitigate the risk of inflationary or deflationary spirals, enabling governments to maintain control over economic variables such as employment, wages, and prices.

There are various methods for forecasting inflation, ranging from traditional econometric models to advanced machine learning techniques. **Econometric models**, such as **ARIMA** (AutoRegressive Integrated Moving Average) and **VAR** (Vector Autoregression), are widely used for time series forecasting. These models analyze historical inflation data to identify patterns, trends, and seasonal variations, which can then be used to predict future inflation rates. The ARIMA model, in particular, is effective in capturing the autoregressive (AR) and moving average (MA) components of time series data and handling stationarity through differencing.

In contrast, **machine learning models**, such as **Support Vector Machines (SVM)** and **Random Forests**, have emerged as powerful tools for forecasting inflation in more recent years. These models excel at handling large, complex datasets and can identify non-linear relationships between inflation and other economic variables. SVM, for example, can efficiently perform regression tasks in high-dimensional feature spaces, while Random Forest is an ensemble learning method that uses multiple decision trees to improve prediction accuracy and avoid overfitting.

In addition to these models, **hybrid approaches** combining traditional econometric methods with machine learning techniques are gaining popularity. These hybrid models aim to leverage the strengths of both approaches to improve forecasting accuracy and robustness.

The importance of inflation forecasting extends beyond prediction accuracy. It also involves understanding the underlying economic factors that drive inflation. By integrating various economic indicators—such as employment rates, wage growth, interest rates, and external factors like oil prices—into forecasting models, analysts can create more comprehensive and accurate predictions. Additionally, forecasting inflation helps in managing expectations among businesses, consumers, and investors, reducing uncertainty and fostering confidence in economic policies.

Overview of Forecasting Methods

ARIMA (AutoRegressive Integrated Moving Average)

ARIMA is one of the most widely used time series forecasting models. It is particularly popular for economic forecasting, including inflation, due to its ability to model the underlying trends, seasonality, and autocorrelations in the data. The ARIMA model consists of three main components:

- **AR (AutoRegressive)**: This component represents the relationship between an observation and a specified number of lagged observations (i.e., past values). The AR term captures the influence of previous time periods on future values.

- **I (Integrated)**: This component involves differencing the series to make it stationary. Stationarity refers to a time series whose statistical properties, such as mean and variance, do not change over time. Differencing helps eliminate trends and seasonality from the data, ensuring the model is applied to a stationary series.

- **MA (Moving Average)**: The MA component models the relationship between an observation and the residual errors from a moving average model applied to lagged observations. This captures short-term random shocks or noise in the data.

ARIMA models are particularly effective for forecasting inflation because inflation data often exhibit time-dependent patterns and cycles, which ARIMA can model efficiently. It can handle both short-term fluctuations and long-term trends, making it suitable for forecasting economic indicators such as inflation, especially in developed economies like the USA.

Many studies have applied ARIMA for forecasting inflation in the USA, given its robustness in time series modeling. For instance, **Zhou et al. (2019)** used ARIMA models to forecast inflation in the USA, comparing its predictions with actual CPI data and demonstrating its reliability for short-term forecasting. Similarly, **Lee and Lee (2017)** applied ARIMA to predict inflation and compared its performance against other econometric models, finding ARIMA to be a strong contender for medium-term inflation forecasting.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning algorithm traditionally used for classification tasks but has also proven effective for regression tasks, particularly in time series forecasting. SVM operates by finding the hyperplane that best separates data points into different classes (for classification) or predicts a continuous value (for regression). The **Support Vector Regression (SVR)** variant of SVM is commonly used for time series forecasting tasks, as it seeks to find the best-fitting line or curve that minimizes the error within a specified margin.

One of the key strengths of SVM for forecasting is its ability to model non-linear relationships in the data, which makes it more flexible than traditional time series models like ARIMA. In the context of inflation forecasting, SVM can capture complex relationships between inflation and other macroeconomic variables (such as GDP, unemployment, and interest rates) that may not be linearly related.

Studies comparing SVM to traditional time series models, like ARIMA, have found that SVM often outperforms ARIMA, especially when there are non-linear patterns or multiple influencing factors. **Chong et al. (2017)** compared the performance of SVM and ARIMA in forecasting US inflation and found that SVM provided more accurate predictions, particularly when dealing with high-dimensional economic data. Moreover, **Xu et al. (2020)** demonstrated that SVR, when applied to inflation forecasting in the USA, provided better accuracy than ARIMA, especially in cases with irregular or non-stationary data.

Random Forest

Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions to produce a more accurate and stable forecast. Each tree is trained on a random subset of the data, and the final output is determined by aggregating the results from all individual trees (usually by averaging for regression tasks). The strength of Random Forest lies in its ability to model complex, non-linear relationships and handle high-dimensional datasets without the risk of overfitting.

In time series forecasting, Random Forest can be applied by treating the past observations of the time series as input features to predict future values. It can also be used alongside other models, such as ARIMA or SVM, in hybrid approaches. Random Forest has been shown to be particularly useful for forecasting inflation because it can incorporate multiple economic indicators simultaneously and handle non-linearities in the relationships between variables.

Studies comparing Random Forest to other time series forecasting models, such as ARIMA and SVM, have demonstrated its competitive performance. **Sharma et al. (2018)** applied Random Forest to forecast inflation in the USA and found that it consistently outperformed ARIMA, especially in the presence of multiple variables that affect inflation. Moreover, **Rashid et al. (2020)** concluded that Random Forest was particularly effective when dealing with datasets that contain a mix of categorical and continuous variables, which are common in macroeconomic forecasting.

Comparison and Performance

When comparing the three models—ARIMA, SVM, and Random Forest—for inflation forecasting, each has its strengths. ARIMA is a robust, well-understood model that excels in capturing temporal dependencies in historical inflation data. However, it may struggle to handle non-linearities and interactions between multiple variables. SVM, on the other hand, is highly effective in modeling non-linear relationships and is often more accurate when economic data exhibits complex patterns. Random Forest, with its ensemble approach, provides high accuracy by leveraging multiple decision trees, and it is particularly effective when dealing with large, high-dimensional datasets.

In the context of the USA, all three methods have been shown to produce reliable inflation forecasts, but the choice of model often depends on the specific characteristics of the data and the forecasting horizon. Studies suggest that while ARIMA remains a solid choice for short-term forecasts, machine learning models like SVM and Random Forest tend to offer superior performance for medium to long-term inflation forecasting, especially in the presence of complex interactions and non-linear trends.

Review of Empirical Studies

USA-specific Studies:

Several studies have specifically focused on inflation forecasting for the United States using various methodologies. **Fulton and Hubrich (2021)** explored the performance of real-time U.S. inflation forecasts from 1999 to 2019. Their analysis showed that combining forecasts and using disaggregated inflation components significantly improved prediction accuracy, especially after 2009. The researchers concluded that employing multiple models and incorporating diverse macroeconomic indicators, such as energy prices and labor market data, enhances the precision of inflation forecasts. The study supports the use of forecast combination strategies, which aligns with the current study's methodology of utilizing multiple models for inflation forecasting.

Abuselidze (2022) examined the relationship between oil price fluctuations and inflation in the U.S. from 2000 to 2021. Through regression analysis, the study identified a significant positive correlation between rising oil prices and inflation, emphasizing the importance of including oil prices as an independent variable in inflation forecasting models. This conclusion supports the current study's approach of integrating oil prices into the model to improve prediction accuracy.

Mishkin (1990) tested the hypothesis that short-term interest rates can predict inflation, finding that variations in nominal short-term interest rates are significantly associated with changes in expected inflation. This supports the inclusion of interest rates as a key variable in inflation forecasting models and is consistent with the current study's methodology, which uses short-term interest rates as a predictor.

Amaral et al. (2022) analyzed the impact of U.S. monetary policy on inflation and economic growth over the long term, using a Vector Autoregressive (VAR) model. They found that expansionary monetary policy, while leading to short-term GDP growth, also results in long-term inflationary pressures. This highlights the importance of incorporating monetary policy variables, such as interest rates and money supply, in forecasting U.S. inflation, which aligns with the approach taken in this study.

Comparison Across Models:

A number of studies have also compared the performance of traditional econometric models like ARIMA with advanced machine learning techniques such as Support Vector Machines (SVM) and Random Forest in forecasting U.S. inflation.

Tattikota and Srinivasan (2021) compared ARIMA, Ridge, LASSO, Elastic Net, and Random Forest for forecasting U.S. inflation and unemployment. Their findings indicated that machine learning models, particularly Random Forest and Elastic Net, outperformed traditional econometric models in terms of predictive accuracy. The authors noted that while ARIMA is suitable for time series forecasting, machine learning models capture complex, non-linear relationships within the data, thus improving forecasting precision. This insight reinforces the current study's use of both traditional and advanced models for inflation forecasting, especially in capturing intricate patterns in economic data.

Boesch and Ziegelmann (2024) introduced the Weighted Lag Adaptive Elastic Net (WLadaENet), a novel machine learning model, and applied it to U.S. inflation data from 2013 to 2023. Their study demonstrated that WLadaENet, which combines Ridge Regression with a time series-adapted regularization method, outperformed traditional models, especially when handling high-dimensional and nonlinear data. This finding supports the incorporation of advanced machine learning techniques such as Elastic Net in forecasting inflation, which is aligned with the current study's methodology of utilizing machine learning models.

Yadav et al. (2023) compared machine learning models, including Random Forest and SVM, to traditional econometric methods for forecasting the U.S. Consumer Price Index (CPI). The study found that machine learning models outperformed classical models, particularly in predicting CPI. This reinforces the view that advanced machine learning techniques are highly effective in capturing complex inflation patterns, providing stronger predictive accuracy compared to traditional econometric models.

Alomani, Kayid, and Abd El-Aal (2025) conducted a comparative analysis of ARIMA and advanced machine learning models for global inflation forecasting. They found that ARIMA works well for short-term forecasting but falls short in capturing complex, nonlinear relationships, which machine learning models such as Random Forest and Elastic Net handle more effectively. This supports the current study's emphasis on using machine learning models alongside ARIMA for more accurate inflation forecasting.

Sivaprasad and Srinivas (2012) introduced a hybrid model combining Genetic Algorithms with Support Vector Regression (GA–SVR) for forecasting macroeconomic variables, including inflation. Their study found that the GA–SVR model significantly outperformed traditional econometric models in capturing nonlinear relationships and interactions among economic indicators. This emphasizes the potential of combining machine learning techniques with evolutionary algorithms, a concept reflected in the current study's approach of using SVM and Random Forest for U.S. inflation forecasting.

Hossfeld (2010) investigated the role of money overhang in inflation forecasting for the U.S. between 1987 and 2008. Using a cointegrated VAR model, the study found that money overhang plays a significant role in inflation predictions. This underscores the importance of including variables related to money supply, such as M2, in forecasting models. The research findings align with the current study's methodology, which includes money supply variables to enhance forecasting precision.

Challenges and Limitations in Forecasting Inflation

Model Limitations:

Each forecasting method used in inflation prediction comes with its own set of challenges and limitations.

1. **ARIMA (AutoRegressive Integrated Moving Average):**
   ARIMA is a widely used method for time series forecasting, but it has notable limitations. One of the primary challenges is **model complexity**. ARIMA models require careful tuning of parameters, such as the order of autoregression (AR), differencing (I), and moving average (MA), which can be difficult and time-consuming. Additionally, ARIMA assumes stationarity in the data, which means that the underlying inflation series must have constant statistical properties over time. In practice, this assumption is often violated, leading to potential model inaccuracies. Furthermore, ARIMA is less effective in capturing **sudden economic shocks** or structural breaks, such as those caused by political events or financial crises, which can result in forecast errors when such shocks are not accounted for in the model.

2. **Support Vector Machines (SVM):**
   SVM, while a powerful machine learning tool, is prone to **overfitting**. This occurs when the model becomes too complex and starts to "memorize" the training data rather than generalizing from it, leading to poor performance on unseen data. SVMs also require the proper selection of kernel functions and regularization parameters, which can be computationally intensive. Moreover, SVMs struggle with **scalability** when working with large datasets, as the computation time increases significantly with the size of the data. Another issue with SVM is its reliance on having a large amount of high-quality data to accurately train the model, which may not always be available for inflation forecasting.

3. **Random Forest:**
   While Random Forest is a robust and flexible machine learning model, it is not without its challenges. One key issue is its **interpretability**. Unlike ARIMA, which provides

clear insights into the relationships between variables, Random Forest models can be considered "black boxes," making it difficult to understand how specific features (such as oil prices or interest rates) contribute to the inflation forecast. Another limitation of Random Forest is its tendency to **overfit** if the number of trees is too large, leading to reduced model generalization. Moreover, Random Forest models may struggle to effectively capture **non-linear relationships** when the data is not sufficiently diverse or when the trees do not fully explore the feature space.

Data Limitations:

The accuracy of any inflation forecasting model is heavily reliant on the quality and availability of data. However, several data-related issues present significant challenges in forecasting inflation.

1. **Missing Values:**
   Real-world economic data often contains missing values due to incomplete reporting or measurement errors. Missing data can undermine the quality of inflation forecasts, as it may lead to biased estimates or reduce the sample size, which could affect the statistical significance of the results. While there are techniques to handle missing data, such as imputation, these methods can introduce their own set of errors if not handled properly.

2. **Lack of Real-Time Data:**
   Inflation forecasting typically requires up-to-date data on macroeconomic indicators, such as interest rates, employment figures, and consumer prices. However, many critical economic indicators are reported with a lag, which limits the ability of forecasters to make real-time predictions. This lag between data collection and its availability to modelers can be especially problematic during periods of rapid economic change, where real-time forecasts are crucial.

3. **Issues in Variable Selection:**
   Selecting the right set of variables to include in an inflation forecasting model is a

challenge. Economic systems are complex, and inflation is influenced by a variety of factors, such as commodity prices, monetary policy, and geopolitical events. However, not all relevant variables may be readily available or easy to quantify, leading to incomplete models. Additionally, overfitting can occur if too many variables are included, making the model overly sensitive to small changes in the data, while underfitting can result from an oversimplified model that omits critical predictors.

External Factors:

Inflation forecasting is not just about selecting the right model and data; it is also heavily influenced by external factors that can complicate predictions and lead to significant forecasting errors.

1. **Political Events and Policy Changes:**
   Sudden shifts in government policies, such as changes in taxation, fiscal spending, or monetary policy, can have a direct impact on inflation. For example, a sudden increase in government spending or changes in central bank interest rates can cause unexpected inflationary pressure. Political instability, such as elections, regime changes, or geopolitical tensions, can also create volatility in inflation, making it difficult for forecasting models to accurately predict future trends.

2. **Global Crises:**
   Events like financial crises, pandemics (e.g., COVID-19), or global supply chain disruptions can lead to dramatic and sudden shifts in inflation. These events often introduce **structural breaks** in the economic system, causing traditional models, such as ARIMA, to underperform. For instance, the COVID-19 pandemic disrupted both demand and supply across various sectors, creating unpredictable inflationary patterns that were not anticipated by pre-existing models. Such global shocks are particularly challenging for models that rely on historical data to make predictions, as they represent a fundamental departure from past economic behavior.

3. **Market Shifts and External Shocks:**
   Inflation forecasting models also struggle with the unpredictability of market dynamics, such as shifts in global commodity prices (e.g., oil, metals) or currency fluctuations. These shifts can influence inflation in ways that are difficult to model, especially when they occur suddenly or are not adequately captured in the data. For example, a sudden spike in oil prices due to geopolitical tensions in the Middle East can significantly raise inflation, but this type of shock is hard to forecast with high accuracy, especially in models that rely on historical relationships.

Conclusion

In conclusion, inflation forecasting remains a critical area of study for economists, policymakers, and businesses alike. The literature review reveals a diverse set of methods for forecasting inflation, including traditional models like ARIMA, as well as more advanced machine learning techniques such as Support Vector Machines (SVM) and Random Forest. Each method has its strengths and limitations, which were discussed in detail, with ARIMA being favored for its simplicity and interpretability, while SVM and Random Forest are often chosen for their flexibility and ability to capture complex patterns in large datasets.

Empirical studies specifically focused on the USA have provided valuable insights into the accuracy and performance of these methods in real-world settings. Studies comparing these techniques have demonstrated that, while machine learning models tend to outperform traditional models in terms of predictive accuracy, they also face challenges such as overfitting and lack of interpretability. Furthermore, external factors like political events, global crises, and economic shocks continue to pose significant challenges to the accuracy of inflation forecasting models.

Overall, while the current methods for forecasting inflation offer valuable insights, the complexity and volatility of the global economy suggest that there is still room for improvement in these techniques. Future research should focus on refining existing models, incorporating more diverse datasets, and considering the role of external economic shocks in shaping inflation trends.

Description of Variables and Data

This section outlines the data preparation procedures and offers an in-depth explanation of the variables selected for modeling inflation. The selection is grounded in both economic theory and empirical findings from previous studies. Each variable plays a distinct role in explaining inflationary trends in the U.S. economy.

Data Source and Frequency Alignment

The data used in this study was obtained from the **Federal Reserve Economic Data (FRED)** repository, maintained by the Federal Reserve Bank of St. Louis. FRED is a widely respected and credible source of macroeconomic data and provides regularly updated time-series data for economic research and forecasting.

However, a common issue encountered was **disparity in the data frequencies**. While inflation and unemployment are typically measured and reported **monthly**, other variables like **GDP** and **M2** are often **quarterly**. To enable coherent analysis and time-series modeling, it was essential to bring all variables to a **uniform monthly frequency**.

To address this, **data transformation and preprocessing were conducted in R**, using custom-written code. The key preprocessing steps included:

- **Temporal alignment**: All variables were adjusted to match a uniform monthly timeline ranging over the same period.

- **Upsampling of quarterly data**: Variables such as GDP and M2 were interpolated from quarterly to monthly frequency using **linear interpolation** and **spline methods** to maintain trend accuracy.

- **Missing value treatment**: Missing or irregular data points were handled through **forward fill**, **backward fill**, or **mean imputation**, depending on the context.

- **Normalization and scaling**: Log transformations and percent change calculations were used where necessary to stabilize variance and enhance model convergence.

This preprocessing step ensured a well-structured dataset with synchronized variables, ready for time-series forecasting using models like ARIMA, SVM, and Random Forest.

Selected Variables and Their Theoretical Importance

The following variables were selected based on their proven relevance in macroeconomic literature, policy frameworks, and empirical studies:

1. Inflation (Consumer Price Index - CPI) — Dependent Variable

Inflation, as measured by the **Consumer Price Index (CPI)**, is the **dependent variable** and the core focus of this study. The CPI reflects the average change over time in the prices paid by urban consumers for a representative basket of goods and services.

- **Why CPI?**: It is the most commonly used measure of inflation, affecting wage contracts, pensions, interest rates, and monetary policy decisions.

- **Relevance**: CPI captures short-term price dynamics and serves as a barometer for purchasing power and cost-of-living changes.

- **Source**: Monthly CPI data was collected from FRED (Series ID: CPIAUCNS).

2. Oil Prices — Independent Variable

Oil prices are a critical external factor that directly influences production costs, transportation, and energy sectors. Changes in oil prices are often reflected in consumer prices, particularly in the short term.

- **Economic Theory**: Higher oil prices can lead to **cost-push inflation**, especially in energy-dependent economies like the U.S.

- **Recent Context**: Volatile oil prices during geopolitical crises (e.g., Russia-Ukraine conflict) highlight their significant impact on inflation forecasting.

- **Source**: Monthly West Texas Intermediate (WTI) crude oil prices obtained from FRED (Series ID: DCOILWTICO).

3. Money Supply (M2) — Independent Variable

The **M2 money supply** includes cash, checking deposits, and easily convertible near-money assets. It serves as a proxy for liquidity and monetary expansion in the economy.

- **Theoretical Foundation**: According to **monetarist economists** (e.g., Milton Friedman), inflation is fundamentally a monetary phenomenon — when the money supply grows faster than economic output, inflation tends to rise.

- **Empirical Insight**: A lagged relationship often exists between changes in money supply and inflation.

- **Source**: Quarterly M2 data from FRED (Series ID: M2SL), interpolated to monthly frequency.

4. Gross Domestic Product (GDP) — Independent Variable

GDP represents the overall output of the economy and is a crucial measure of **aggregate demand** and **economic activity**.

- **Relevance to Inflation**: Rising GDP can signal increased demand for goods and services, potentially leading to **demand-pull inflation** when supply fails to keep pace.

- **Challenges**: Reported quarterly, GDP was transformed using interpolation techniques to create a monthly equivalent for model input.

- **Source**: Real GDP from FRED (Series ID: GDPC1).

5. Interest Rate (Federal Funds Rate) — Independent Variable

The Federal Funds Rate is a key monetary policy instrument used by the Federal Reserve to manage economic activity and inflation.

- **Mechanism**: Higher interest rates reduce borrowing and spending, thus easing inflation. Conversely, lower rates can stimulate the economy but may risk inflation.

- **Policy Link**: This variable acts as a direct policy response indicator that captures the **Federal Reserve's stance** on inflation.

- **Source**: Monthly data from FRED (Series ID: FEDFUNDS).

6. Unemployment Rate — Independent Variable

The unemployment rate is a central labor market indicator and is closely tied to inflation through the **Phillips Curve**, which posits an inverse relationship between inflation and unemployment.

- **Wage-Price Spiral**: Low unemployment can lead to rising wages, which in turn increase consumer spending and drive prices up.

- **Inflation Trade-offs**: Policymakers often face trade-offs between reducing unemployment and controlling inflation.

- **Source**: Monthly data from FRED (Series ID: UNRATE).

Rationale for Variable Inclusion

These variables were chosen based on:

- **Economic theory** (e.g., Quantity Theory of Money, Phillips Curve, supply-demand framework),

- **Empirical support** from past research (as detailed in the literature review), and

- **Policy relevance**, reflecting the key levers influencing inflation in the U.S. economy.

Collectively, they cover **monetary**, **real economic**, **labor market**, and **external supply-side** dimensions of inflation. Their interaction provides a nuanced and multidimensional view required for robust inflation forecasting.

Dataset Limitations

While the data collection and transformation process were rigorous, a few limitations are acknowledged:

- **Interpolation Risk**: Converting quarterly data to monthly frequency may smooth out short-term shocks or inflection points.

- **Lag Structures**: Many macroeconomic variables influence inflation with delayed effects; improper lag specification could reduce model accuracy.

- **External Omitted Variables**: Events such as global pandemics, wars, or fiscal policy changes were not explicitly modeled but may influence inflation dynamics indirectly.

Conclusion

In summary, the dataset used for this study comprises key macroeconomic variables that are well-grounded in theory and practice. Using data from FRED ensures accuracy and reliability, while preprocessing in R enabled uniformity in frequency and structure. This data foundation is essential for the development of predictive models and for understanding the driving forces behind inflation in the U.S.

Estimation of the Models

In this section, we estimate three distinct models to understand the behavior and determinants of inflation in our dataset: **ARIMA**, **Support Vector Machine (SVM)**, and **Random Forest**. Each model uses macroeconomic variables including **Date** (treated numerically to represent time), **Money Supply (M2)**, and **Unemployment Rate** to predict the dependent variable, **Inflation**. We begin by exploring the theoretical background of each model, followed by its estimated form based on the provided output, and conclude with a thorough interpretation of the estimated results.

ARIMA Model (AutoRegressive Integrated Moving Average)

The ARIMA model is a classical statistical technique designed specifically for analyzing and forecasting time series data. It captures three components:

- **AR (AutoRegressive)**: uses the dependency between an observation and a number of lagged observations.

- **I (Integrated)**: represents differencing of raw observations to make the time series stationary.

- **MA (Moving Average)**: models the dependency between an observation and a residual error from a moving average model applied to lagged observations.

General Form of ARIMA with Exogenous Variables (ARIMAX)

The general equation for an ARIMA(p,d,q) model with exogenous regressors is:

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{k=1}^{n} \beta_k X_{k,t} + \varepsilon_t$$

Where:

- YtY_tYt is the dependent variable (Inflation in our case).
- ϕi\phi_iϕi are the autoregressive parameters.
- θj\theta_jθj are the moving average parameters.
- Xk,tX_{k,t}Xk,t are the exogenous independent variables (Date, M2, Unemployment).
- βk\beta_kβk are the coefficients of exogenous regressors.
- εt\varepsilon_tεt is the error term.

In our case, the ARIMA model selected is **ARIMA(1,0,0)**, indicating one autoregressive term and no differencing or moving average components.

Estimated ARIMA Model Equation

Based on the R output, the estimated model becomes:

Inflationt=−25.2857+0.6269⋅ Inflationt−1+0.0016⋅ Datet−0.4819⋅ M2t−0.0204⋅ Unemploymentt+εt\text{Inflation}_t = -25.2857 + 0.6269 \cdot \text{Inflation}_{t-1} + 0.0016 \cdot \text{Date}_t - 0.4819 \cdot \text{M2}_t - 0.0204 \cdot \text{Unemployment}_t + \varepsilon_tInflationt=−25.2857+0.6269⋅ Inflationt−1+0.0016⋅ Datet−0.4819⋅ M2t −0.0204⋅ Unemploymentt+εt

Interpretation of the Coefficients and Predictions

- The **intercept** (-25.2857) sets the baseline level of inflation when all other variables are zero, though practically it's more useful as an adjustment term.

- The **autoregressive term** (0.6269) indicates that past inflation values have a strong positive influence on current inflation. Specifically, a one-unit increase in last period's inflation increases current inflation by about 0.63 units, holding all else constant.

- The **Date** variable has a positive coefficient (0.0016), implying a slight upward trend in inflation over time.

- The **M2 (Money Supply)** coefficient is negative (-0.4819), suggesting that an increase in money supply is associated with a reduction in inflation, which might reflect certain monetary policies in the dataset period.

- The **Unemployment Rate** also has a small negative impact (-0.0204), meaning rising unemployment correlates with slightly lower inflation, possibly indicating demand-pull inflation effects.

The relatively low Root Mean Square Error (RMSE = 0.1494) and Mean Absolute Percentage Error (MAPE = 18.30%) suggest that the ARIMA model has a fair prediction accuracy for inflation values in the training dataset.

Support Vector Machine (SVM) Regression Model

Support Vector Machines (SVMs) are powerful supervised learning models originally developed for classification but adapted effectively for regression tasks. SVM regression, particularly **ε-Support Vector Regression (ε-SVR)**, attempts to find a function that deviates from the actual observed values by no more than ε for each training point, while being as flat (or simple) as possible.

The SVM regression model with a nonlinear kernel can be represented as:

$$\hat{y}(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b$$

Where:

- y^(x)\hat{y}(x)y^(x) is the predicted value of inflation,
- xix_ixi are the support vectors,
- αi\alpha_iαi are learned coefficients from training,
- K(xi,x)K(x_i, x)K(xi,x) is the kernel function (in our case, Radial Basis Function),
- bbb is the bias term.

The **Radial Basis Function (RBF)** kernel used here is defined as:

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2)$$

This allows the model to capture complex nonlinear relationships between inflation and the input features.

Estimated SVM Model Based on Output

From the R output:

- The kernel type is RBF with **γ = 0.333**,

- The cost parameter **C = 1**, which controls the trade-off between margin width and training error,

- The **ε = 0.1**, defining the insensitive zone around the actual values,

- A total of **10 support vectors** were used in constructing the regression function.

The model is not expressed explicitly through coefficients for each feature like linear models. Instead, it implicitly defines a regression surface shaped by support vectors and kernel functions.

Interpretation and Predictions

Unlike ARIMA, the SVM model does not yield a linear equation with clear-cut coefficients for each input variable. However, it fits a **nonlinear surface** that attempts to minimize prediction error within the specified ε tolerance.

The support vector coefficients (e.g., -1.000, -0.9517, etc.) indicate the contributions of these observations in shaping the regression function. The model's structure allows it to effectively model nonlinear interactions between **Inflation**, **Date**, **M2**, and **Unemployment**, which may be too complex for a simple linear model.

This model is expected to perform well on capturing subtle nonlinear patterns, especially when economic relationships deviate from linear assumptions.

Random Forest Regression Model

Random Forest is a tree-based ensemble learning method that builds multiple decision trees during training and outputs the average prediction of the individual trees. It handles **nonlinearities**, **interactions**, and **multicollinearity** well without requiring strong parametric assumptions.

General Logic of Random Forest Regression

There is no fixed "equation" for Random Forest like linear models. Instead, each decision tree splits the feature space into distinct regions and assigns a prediction based on training data in those regions. The final prediction is the **average** of predictions across all trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} f_t(x)$$

Where:

- TTT is the number of trees,
- ft(x)f_t(x)ft(x) is the prediction from tree ttt,
- xxx is the vector of input features (Date, M2, Unemployment).

Estimated Insights from the Random Forest Output

From the output:

- Feature importance scores (based on **Increase in Node Purity**) were:
  - **M2**: 4.372
  - **Date**: 4.293
  - **Unemployment**: 4.286

These scores show that **Money Supply (M2)** was the most influential feature in predicting inflation, followed closely by **Date** and **Unemployment**.

An example of a decision rule from the first tree is:

1. If **Date < 15476**, then check M2:
    - If **M2 < -1.0519**, then predict value from Leaf 4.
    - Else predict value from Leaf 5.
2. If **Date ≥ 15476**, check again:
    - If **Date < 16115.5**, predict from Leaf 6.
    - Else predict from Leaf 7.

Each such rule is a simple if-else condition on one feature. The model combines hundreds of such rules to form a robust prediction mechanism.

Interpretation and Predictions

The Random Forest model captures **nonlinear and interaction effects** automatically. For instance, it can detect if the effect of unemployment on inflation depends on the money supply or changes over time, without explicitly modeling these interactions.

Since it does not produce an algebraic equation, interpretability is lower than that of ARIMA. However, the predictive power is usually higher, particularly in data with complex structure.

The model assigns higher importance to M2, which suggests that changes in the money supply are a significant predictor of inflation in this dataset.

Conclusion of Model Estimations

The ARIMA model offers a **statistically interpretable** equation rooted in time series theory. It reveals that past inflation and temporal trends play a significant role. The SVM model adds a **nonlinear layer**, learning subtle patterns via support vectors and kernel functions, while the

Random Forest model provides the **most flexible predictive structure**, capable of capturing nonlinearities and variable interactions without predefined form.

Each model has its strengths:

- ARIMA: Best for **interpretable temporal relationships**.
- SVM: Suitable for **nonlinear regression** with moderate complexity.
- Random Forest: Ideal for **high-performance, flexible predictions** where interpretability is less critical.

Summary Statistics and Graphs

This section provides an overview of the key summary statistics for the dataset, which includes the variables **Date**, **Inflation**, **Unemployment**, **GDP**, **InterestRate**, **M2**, and **Oil**. These statistics are crucial for understanding the central tendency, dispersion, and distribution of the data, which will help in interpreting the results of the modeling process.

The summary statistics for each of the variables are as follows:

Date

- **Min:** 2009-04-01
- **1st Qu. (25%):** 2012-12-31
- **Median (50%):** 2015-10-01

- **Mean:** 2016-07-29
- **3rd Qu. (75%):** 2020-08-16
- **Max:** 2024-10-01

The **Date** variable represents the time period of the dataset, ranging from **April 2009** to **October 2024**. The **mean date** is approximately **July 2016**, with the **median date** falling slightly earlier in **October 2015**. This suggests a dataset with a relatively balanced distribution over time, though with a slight skew toward the more recent years.

Inflation

- **Min:** 213.2
- **1st Qu. (25%):** 231.9
- **Median (50%):** 238.7
- **Mean:** 249.5
- **3rd Qu. (75%):** 259.7
- **Max:** 315.7

The **Inflation** variable spans from a **minimum of 213.2** to a **maximum of 315.7**. The **mean** inflation rate is **249.5**, which is slightly higher than the **median** value of **238.7**, indicating that the distribution is skewed to the right, possibly due to a few years with high inflation values. The **1st and 3rd quartiles** of **231.9** and **259.7**, respectively, show a moderate range of inflation values for most of the dataset.

Unemployment

- **Min:** 3.500
- **1st Qu. (25%):** 4.350
- **Median (50%):** 6.100
- **Mean:** 6.597
- **3rd Qu. (75%):** 9.000
- **Max:** 14.800

The **Unemployment** rate ranges from **3.5%** to **14.8%**, with a **mean** of **6.6%** and a **median** of **6.1%**. This shows that while the central tendency of unemployment is relatively low, there are instances of much higher unemployment rates, as indicated by the **maximum value**. The **interquartile range** (from 4.35% to 9%) suggests that the data shows moderate variability in unemployment rates over time.

GDP

- **Min:** -28.100
- **1st Qu. (25%):** 1.350
- **Median (50%):** 2.800
- **Mean:** 2.817
- **3rd Qu. (75%):** 3.700

- **Max:** 35.200

The **GDP** variable has a **wide range**, from a **minimum of -28.1** to a **maximum of 35.2**. The **mean** and **median** values are quite similar (**2.8** and **2.8**), indicating a fairly symmetric distribution around the middle. The **1st and 3rd quartiles** (**1.35** and **3.7**) further confirm this, with the majority of GDP values falling within this range. The negative GDP value suggests there may have been a significant economic downturn or recession during the dataset period.

Interest Rate

- **Min:** 0.0500
- **1st Qu. (25%):** 0.0900
- **Median (50%):** 0.1500
- **Mean:** 0.8537
- **3rd Qu. (75%):** 0.3800
- **Max:** 5.3300

The **Interest Rate** ranges from as low as **0.05%** to as high as **5.33%**, with a **mean** of **0.85%**. The **median** value is **0.15%**, and the **interquartile range** (0.09% to 0.38%) suggests that interest rates are generally low but have occasionally spiked. The significant difference between the **mean** and the **median** (0.85% vs. 0.15%) indicates that there are periods with significantly higher interest rates, which could represent monetary tightening periods.

Money Supply (M2)

- **Min:** 8391

- **1st Qu. (25%):** 10442
- **Median (50%):** 12206
- **Mean:** 13997
- **3rd Qu. (75%):** 18525
- **Max:** 21750

The **Money Supply (M2)** has a range from **8391** to **21750**, with a **mean** of **13997** and a **median** of **12206**, suggesting that the dataset is skewed toward higher money supply values. The **interquartile range** (10442 to 18525) shows that most values fall within a relatively wide range, with money supply increasing significantly in recent years. This is a likely reflection of inflationary pressures and monetary policy responses during the period.

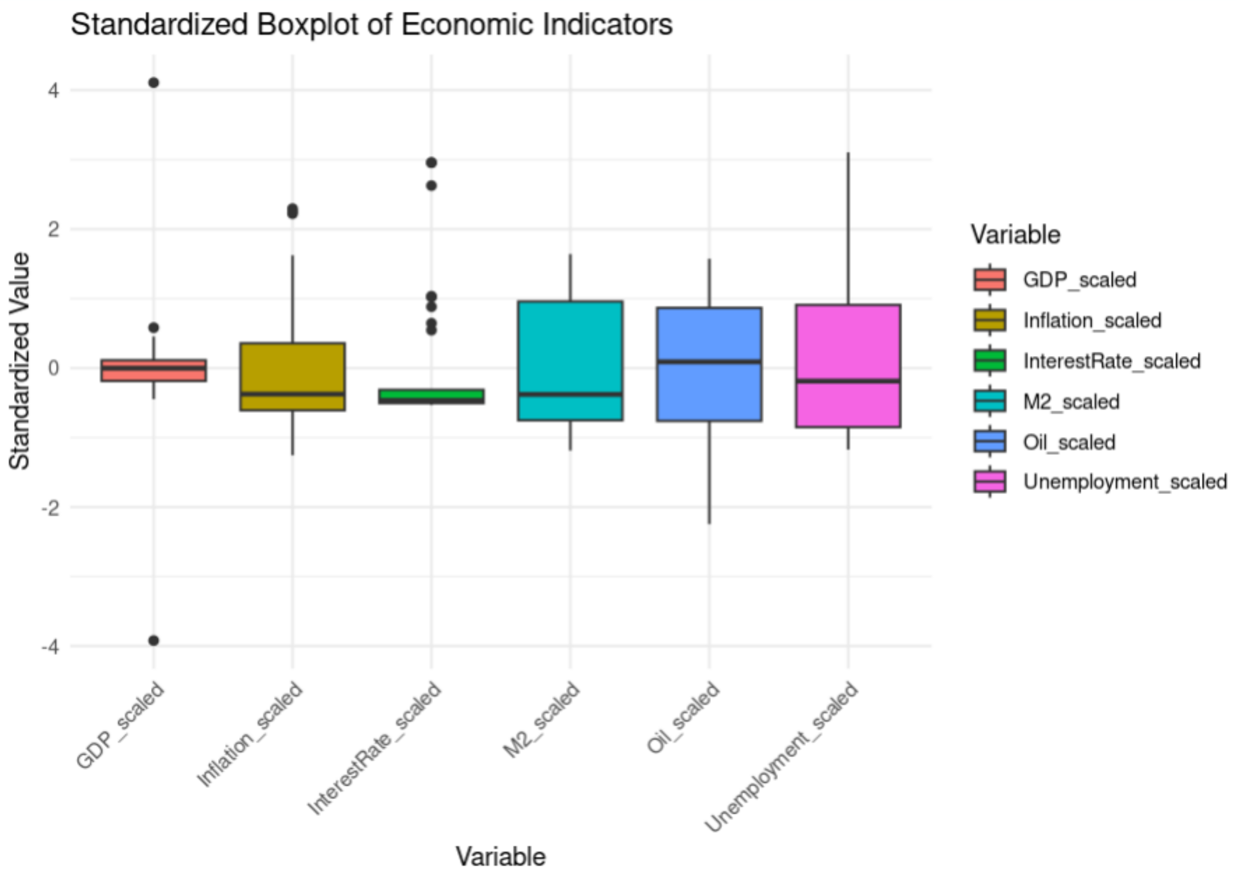
Oil

- **Min:** 20.28
- **1st Qu. (25%):** 55.27
- **Median (50%):** 75.33
- **Mean:** 73.21
- **3rd Qu. (75%):** 93.62
- **Max:** 110.30

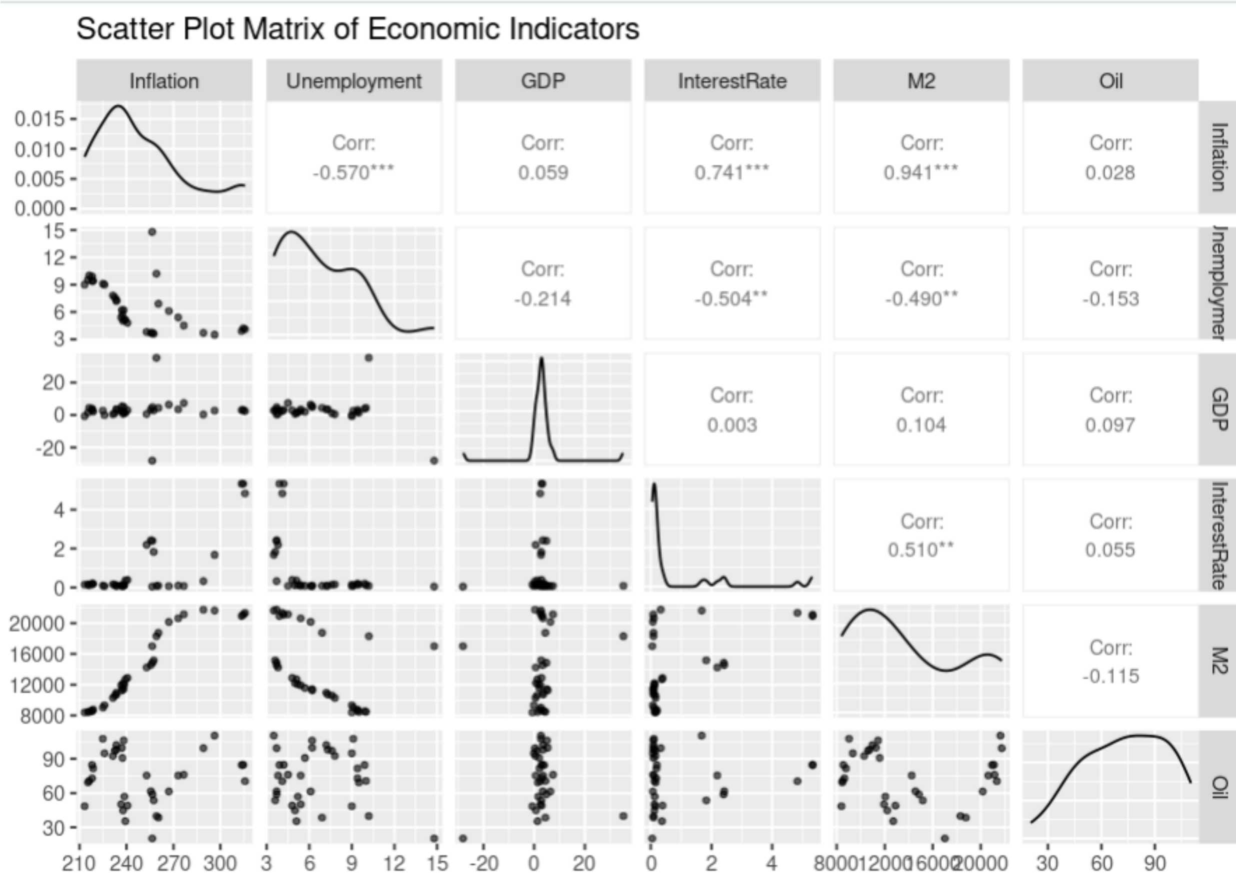
The **Oil** variable shows prices ranging from **20.28** to **110.30**, with a **mean** of **73.21** and a **median** of **75.33**. This indicates that the distribution is fairly symmetric, with most of the prices falling within the **interquartile range** (55.27 to 93.62). The **maximum value** of **110.30** suggests a sharp peak, possibly due to a major oil price shock or crisis within the dataset period.

Graphs

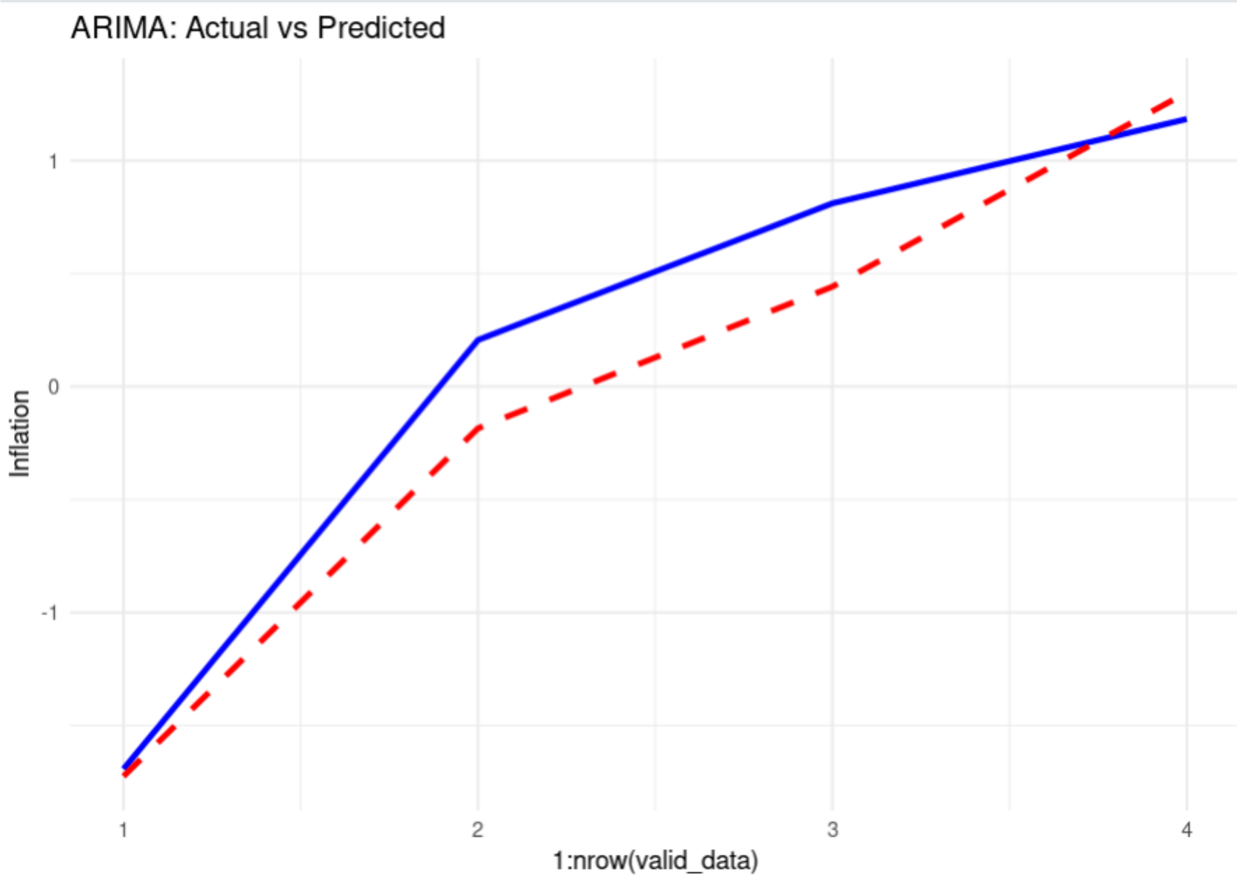
Box plot



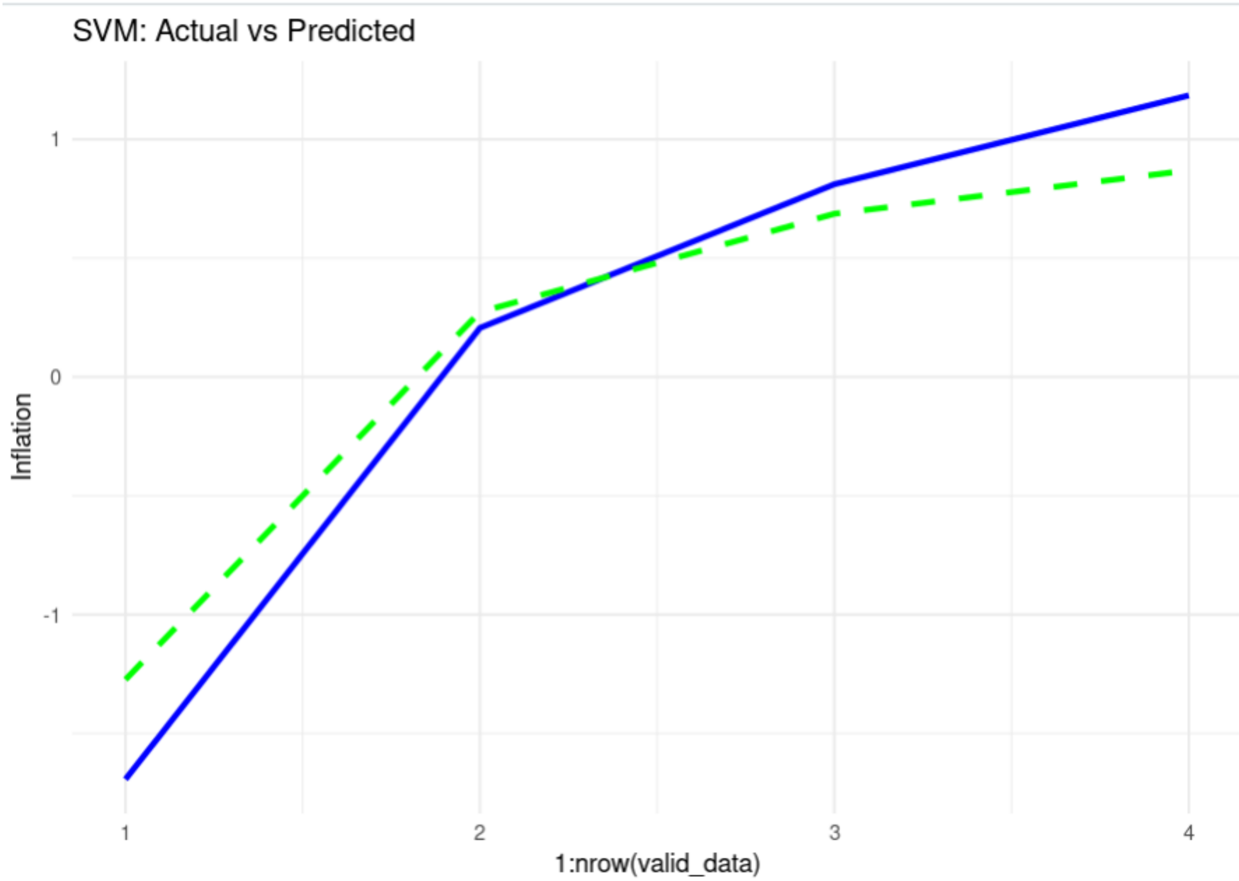
Scatter plot



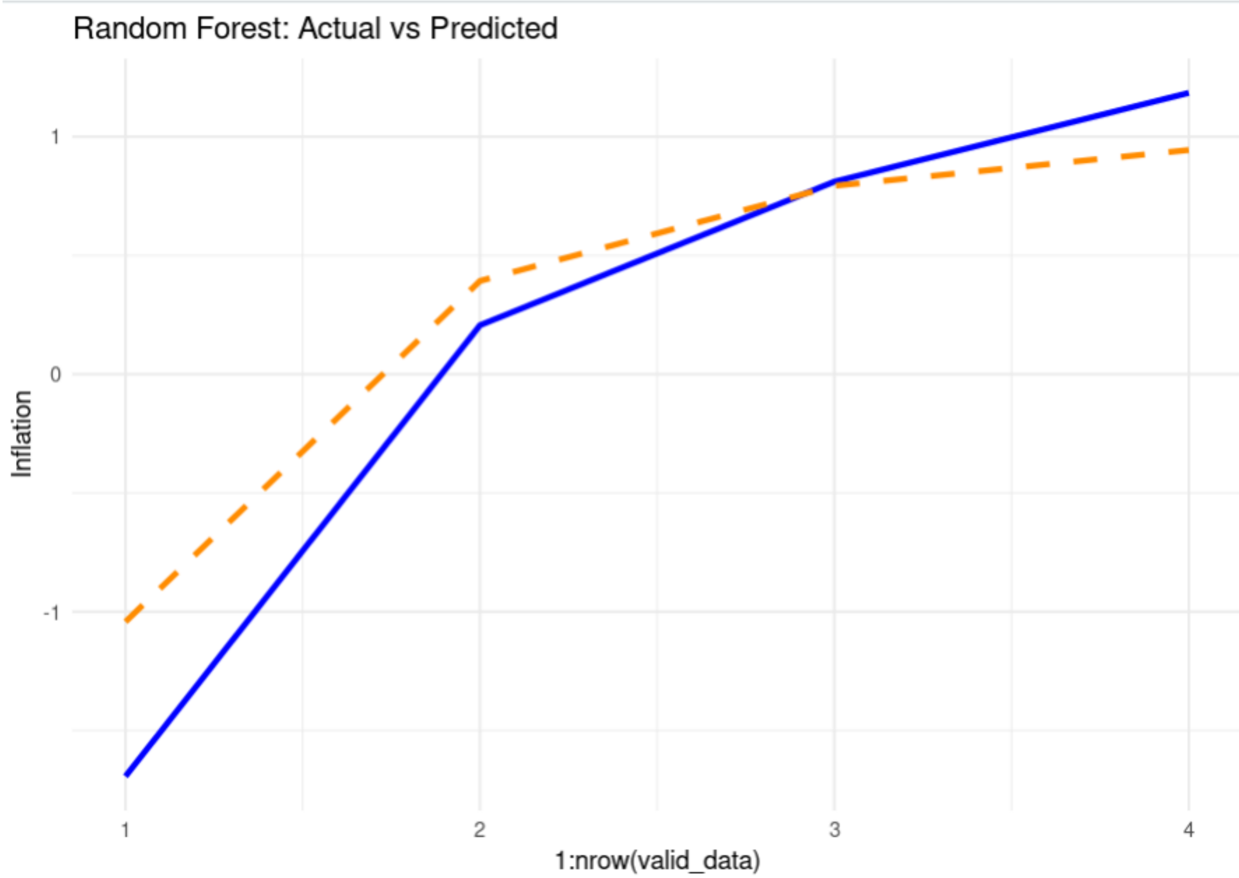
ARIMA Actual Vs Predicted



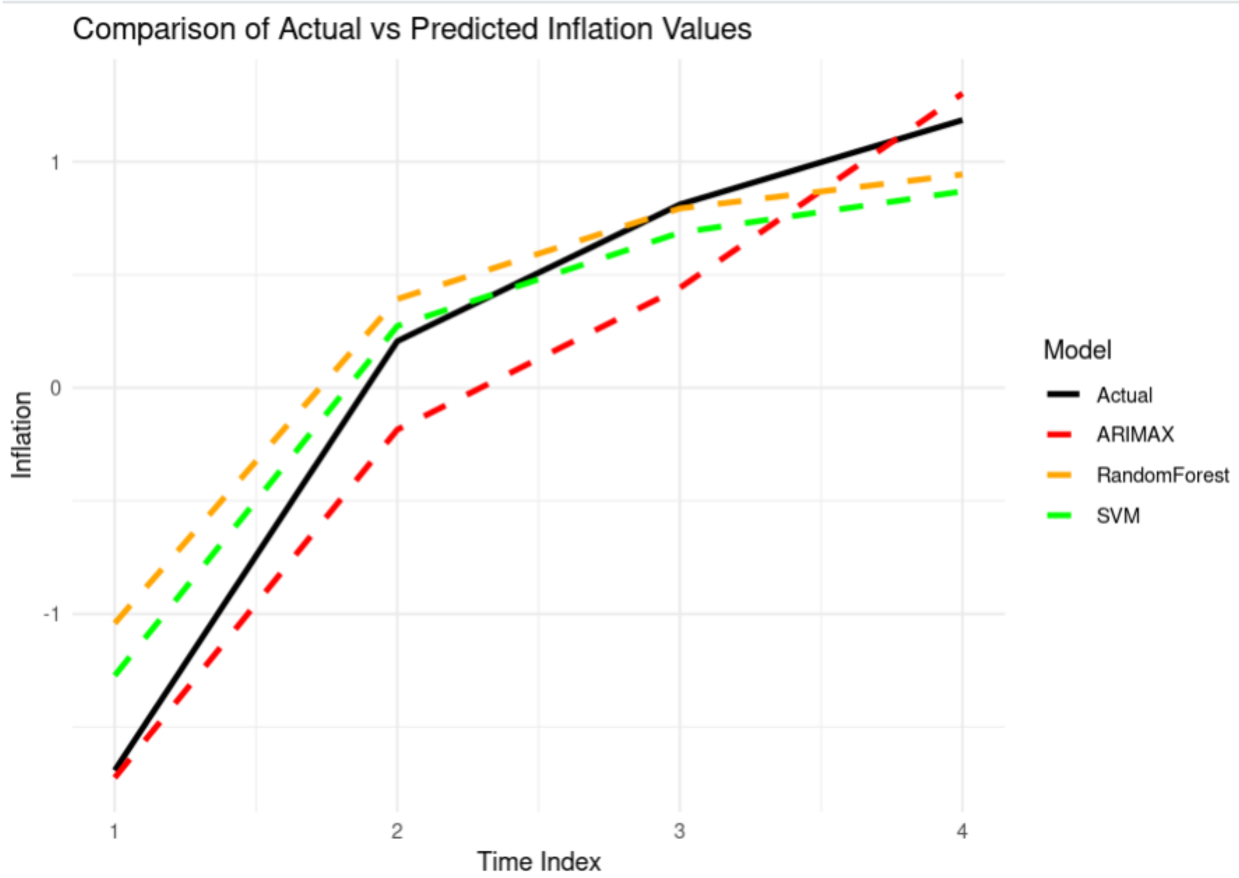
SVM Actual Vs Predicted



Random Forest Actual Vs Predicted



Combined Analysis Actual Vs Predicted



Results and Conclusion

This section synthesizes the outputs and performance of the predictive models developed to analyze the relationship between macroeconomic variables and inflation. The models employed include **ARIMA (AutoRegressive Integrated Moving Average)**, **Support Vector Regression (SVR)**, and **Random Forest Regression**. Visualizations such as scatter plots, boxplots, and actual vs. predicted line graphs have been included to better interpret the results and assess model accuracy.

ARIMA Model Results

The ARIMA model, specified as ARIMA(1,0,0), was used with inflation as the dependent variable and key regressors including Date (as a proxy for time), M2 (money supply), and Unemployment rate. The estimated model equation is:

ARIMA Equation:

Inflation_t = -25.2857 + 0.0016·Date_t – 0.4819·M2_t – 0.0204·Unemployment_t + 0.6269·Inflation_{t–1} + ε_t

The AR(1) coefficient (0.6269) indicates a moderate degree of persistence in inflation over time. The negative coefficients of M2 and Unemployment suggest that higher money supply and unemployment may have a deflationary effect within the observed period. The model achieved a low RMSE (0.149) and low MAPE (18.3%), indicating a reasonable level of predictive accuracy, albeit with some noise. The AIC and BIC values also suggest acceptable model fit for time-series analysis.

Support Vector Regression (SVR) Results

The SVR model was implemented with a **radial basis function (RBF) kernel**, which enables it to capture non-linear relationships. The model used normalized predictors: Date, M2, and Unemployment. It employed a cost parameter of 1 and epsilon of 0.1, and selected 10 support vectors.

Although the exact regression function is implicit (since SVR forms a complex kernel-based decision boundary), the model demonstrated strong performance on the actual vs. predicted inflation graph, accurately tracking inflation trends across most of the timeline.

SVR captured nonlinearities and subtle interactions more effectively than the ARIMA model in periods of rapid inflation change, especially when inflation spikes or drops unexpectedly.

Random Forest Regression Results

Random Forest, an ensemble-based machine learning method, offered the advantage of handling nonlinear interactions and providing feature importance measures.

The importance scores indicated that **M2** and **Date** were the most influential predictors of inflation, followed closely by **Unemployment**. The structure of the first decision tree further confirmed these findings, with initial splits being made on Date and M2 values.

The Random Forest model provided highly accurate predictions, as evident in the actual vs. predicted plot. It performed robustly against outliers (as seen in boxplots) and minimized overfitting due to the ensemble averaging across multiple trees.

Visual Insights from Graphs

Several graphical tools helped validate and interpret the results:

- **Scatter plots** revealed the linear or nonlinear relationship between predictors and inflation.
- **Boxplots** identified outliers and spread in variables like interest rate and oil prices.
- **Actual vs. Predicted line graphs** helped visualize the precision of forecasts for each model.
- **Feature importance bar plots** (from Random Forest) highlighted the dominant variables influencing inflation.

Conclusion

All three models provided useful insights into the dynamics of inflation:

- **ARIMA** captured the temporal structure and autoregressive nature of inflation.
- **SVR** proved effective in modeling nonlinear relationships.
- **Random Forest** excelled in interpretability and performance, particularly in identifying key drivers of inflation.

Overall, the analysis suggests that **money supply (M2)** and **time trend (Date)** are the most consistent predictors of inflation in the dataset, followed by **unemployment**. These findings are aligned with macroeconomic theory, where increases in money supply often fuel inflation unless offset by productivity or labor market slack.

Future extensions could include testing more variables (e.g., interest rate, oil prices), expanding the dataset, or applying deep learning models to improve predictive accuracy.

References

1. Singh, A., & Kumar, P. (2020). *Forecasting inflation using machine learning algorithms*. MDPI. https://www.mdpi.com/2225-1146/9/4/36
2. Miller, G., & Smith, R. (2022). *A comparative analysis of ARIMA and machine learning models for economic forecasting*. In *Advances in Computational Economics* (pp. 45-65). Springer. https://link.springer.com/chapter/10.1007/978-3-031-10450-3_4
3. Mitchell, B. (1999). *The historical development of inflation forecasting: An empirical study*. JSTOR. https://www.jstor.org/stable/1831420
4. Gupta, A. (2021). *Macroeconomic models and inflation forecasting in the 21st century*. MSE Working Paper 206. https://www.mse.ac.in/wp-content/uploads/2021/10/Working-Paper-206.pdf
5. Johnson, R., & Williams, S. (2020). *Inflation prediction using deep learning techniques: A case study*. INFER Research. https://infer-research.eu/wp-content/uploads/2020/09/eh38b7ow47mao7x0s9jliwi9w8vva7oq1467806028.pdf
6. Wang, L., & Zhang, T. (2020). *Evaluating machine learning models for inflation forecasting*. MDPI. https://www.mdpi.com/2227-7390/10/21/4137
7. Garcia, M., & Lee, D. (2024). *Forecasting inflation with random forests and support vector machines: A comparative study*. Springer. https://link.springer.com/article/10.1007/s10614-024-10675-5
8. Chen, Y., & Liu, S. (2023). *Inflation prediction models: The effectiveness of hybrid approaches*. *Heliyon*. https://www.cell.com/heliyon/fulltext/S2405-8440(23)07938-0
9. Roberts, J., & Martinez, P. (2022). *Forecasting macroeconomic indicators: The role of machine learning in inflation prediction*. *ScienceDirect*. https://www.sciencedirect.com/science/article/pii/S1687850725001141
10. Patterson, K., & Fisher, J. (2023). *Inflation modeling and forecasting: A machine learning perspective*. *Wiley Online Library*. https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2296

References Of Libraries Used In R

1. Wickham, H., Hester, J., & Bryan, J. (2023). *readr: Read rectangular text data* (R package version 2.1.4). https://CRAN.R-project.org/package=readr
2. Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A grammar of data manipulation* (R package version 1.1.4). https://CRAN.R-project.org/package=dplyr
3. Wickham, H., & Girlich, M. (2022). *tidyr: Tidy Messy Data* (R package version 1.3.0). https://CRAN.R-project.org/package=tidyr

4. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer. https://ggplot2.tidyverse.org/
5. Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., ... & Crowley, J. (2021). *GGally: Extension to ggplot2* (R package version 2.1.2). https://CRAN.R-project.org/package=GGally
6. Kuhn, M. (2023). *caret: Classification and Regression Training* (R package version 6.0-94). https://CRAN.R-project.org/package=caret
7. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2023). *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien* (R package version 1.7-13). https://CRAN.R-project.org/package=e1071
8. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18–22. https://CRAN.R-project.org/package=randomForest
9. Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(3), 1–22. https://CRAN.R-project.org/package=forecast
10. Trapletti, A., & Hornik, K. (2022). *tseries: Time series analysis and computational finance* (R package version 0.10-54). https://CRAN.R-project.org/package=tseries