Forecasting Inflation in Pakistan Using Advanced Statistical and Time Series Models

By

Muneeb Lone 23i-2623
Rida Zubair 23i-2590

Supervisor
Hazber Samson

Department of Data Science
National University of Computer and Emerging Sciences
Islamabad, Pakistan

Contents

1 Project Title and Type 4

2 Project Introduction 5

2.1 What is Inflation? 5

2.2 Inflation in Pakistan: A Historical Overview 5

2.3 Objective of This Study 5

3 Literature Review 7

3.1 Macroeconomic Determinants of Inflation 7

3.2 Models Used for Inflation Forecasting 7

3.3 Why the Variables Were Chosen 8

3.4 Data Sources 8

3.5 9

4 Variable Selection 10

4.1 Initial Selection of Variables 10

4.2 Data Collection and Refinement Process 10

4.3 Statistical Analysis and Identification of Significant Variables 10

5 Justification for Variable Selection 11

6 Statistical Summary and Insights on Key Variables 14

7 Key Observations and Insights 15

8 Box and Whisker Plots 17

8.1 Boxplot of Variables with Larger Values 17

8.2 Boxplot of All Variables 17

8.3 Key Insights 18

9 Scatter Plot 19

9.1 Interpretation of Pairplot 19

9.2 Conclusion 21

10 Models 22

10.1 1. ARIMA (Autoregressive Integrated Moving Average) Model . . . 22

10.2 2. Ridge Regression 22

10.3 3. LASSO (Least Absolute Shrinkage and Selection Operator) Regression 23

10.4 4. Elastic Net Regression 24

10.5 Comparison and Final Model Selection 24

11 Comparison and Final Model Selection 26

11.1 Comparison of Models 26

11.2 ARIMA Model 26

11.3 Ridge Regression 26

11.4 LASSO Regression 27

11.5 Elastic Net Regression 27

11.6 Comparison of Models 27

12 Results and Discussion 29

12.1 Key Insights 29

12.2 Conclusion 29

13 References 31

1 Project Title and Type

Title: Forecasting Inflation in Pakistan Using Advanced Statistical and Time Series Models

Project Type: Applied Research – Quantitative, Predictive Modeling, and Time Series Analysis

2 Project Introduction

2.1 What is Inflation?

Inflation is the sustained increase in the general price level of goods and services in an economy over time. As prices rise, the purchasing power of money falls, leading to various economic implications. In general, a moderate level of inflation is considered beneficial for economic growth, indicating increased demand and investment. However, when inflation becomes excessive or unpredictable, it can erode purchasing power, destabilize financial systems, and negatively impact the economy.

Inflation is influenced by several macroeconomic factors, including demand and supply dynamics, exchange rates, interest rates, and government policies. For countries like Pakistan, inflation is a persistent issue, often fluctuating due to external shocks, structural inefficiencies, and changing global economic conditions. Consequently, accurately forecasting inflation is vital for sound economic management and policy formulation.

2.2 Inflation in Pakistan: A Historical Overview

Over the past five decades, Pakistan has witnessed significant inflationary fluctuations. From the 1970s through the early 2000s, inflation rates were subject to wide swings, driven by factors such as oil price shocks, political instability, and fiscal imbalances. In recent years, inflation in Pakistan has remained volatile, frequently reaching double digits, especially during periods of economic crisis. The inflationary trend has been compounded by rising fuel prices, currency devaluation, and an expanding trade deficit.

The impact of inflation in Pakistan is felt across all sectors, with consequences for income distribution, poverty levels, and economic stability. These persistent inflationary pressures underscore the importance of developing accurate forecasting models to aid policymakers in managing inflation and implementing timely interventions.

2.3 Objective of This Study

The primary objective of this study is to forecast inflation in Pakistan using a combination of traditional econometric models and modern machine learning techniques. The paper specifically aims to assess the effectiveness of ARIMA, Ridge Regression, LASSO, and ElasticNet Regression models in predicting inflation trends

based on key macroeconomic variables, including GDP growth, exchange rates, money supply, and interest rates.

By analyzing a 50-year dataset from 1973 to 2023, this study seeks to identify which model offers the most accurate predictions for inflation in Pakistan and determine the significant predictors that can best explain inflationary trends.

3 Literature Review

Inflation forecasting is a critical area of study in economic research due to its pivotal role in shaping monetary policies and economic strategies. In our research, we focused on the inflation forecasting models applicable to Pakistan, leveraging a blend of traditional econometric approaches and advanced machine learning techniques. The objective was to identify the key macroeconomic variables that influence inflation in Pakistan and assess the effectiveness of various models in predicting inflation trends.

3.1 Macroeconomic Determinants of Inflation

Our selection of variables was based on the extensive body of research that highlights the determinants of inflation in Pakistan. Studies like ”Determinants of Inflation in Pakistan: A Comprehensive Analysis of Macroeconomic Variables (1991–2022)” emphasize the significance of variables such as imports, GDP growth rate, GDP per capita, military expenditure, population growth, total debt service, and unemployment in influencing inflation. These variables were carefully chosen due to their strong empirical evidence in explaining inflationary trends in Pakistan’s economy.

Further support for this selection is found in ”Exploring Macroeconomic Determinants of Inflation in Pakistan” (Jesciences, 2022), which identifies fiscal deficit, interest rates, and exchange rates as significant determinants. Our study considers these variables to ensure a comprehensive model that reflects the diverse influences on inflation. Additionally, ”Forecasting Inflation through Econometrics Models: An Empirical Study on Pakistani Data” suggests that money supply and fiscal deficit are crucial for forecasting inflation, particularly in Pakistan’s context where fiscal and monetary policies play a dominant role.

3.2 Models Used for Inflation Forecasting

In our study, we utilized both traditional time-series models and machine learning approaches to forecast inflation. ARIMA (Auto-Regressive Integrated Moving Average), which has been widely used in previous studies such as ”Forecasting Inflation, Exchange Rate, and GDP using ANN and ARIMA Models”, was one of the models we applied. ARIMA is effective in capturing the temporal dependencies inherent in inflation data. This model has been successful in many inflation forecasting studies, especially when dealing with time-series data with trends and seasonality. Furthermore, we incorporated machine learning models such as

LASSO (Least Absolute Shrinkage and Selection Operator), Ridge regression, and ElasticNet, which have been demonstrated to outperform traditional methods in complex, data-rich environments. For example, "Inflation Forecasting for Pakistan in a Data-rich Environment" used LASSO and Dynamic Factor Models (DFM) to predict inflation, comparing them with more traditional econometric methods. This study showed that machine learning models, particularly LASSO, offer superior forecasting capabilities when handling large datasets with multiple correlated predictors, a challenge frequently encountered in inflation forecasting. Additionally, "Evaluating the Performance of Inflation Forecasting Models of Pakistan" evaluated several models, both econometric and machine learning, for forecasting inflation. This comparison underscored the value of machine learning methods in improving predictive accuracy, especially for inflation forecasting in Pakistan's volatile economic environment.

3.3 Why the Variables Were Chosen

The variables selected for our model were chosen based on their historical and theoretical relevance to inflation in Pakistan. As established in previous studies like "Exploring Macroeconomic Determinants of Inflation in Pakistan" and "Forecasting Inflation through Econometrics Models: An Empirical Study on Pakistani Data", GDP growth, exchange rates, money supply, and trade balance have been identified as fundamental drivers of inflation. These variables are crucial for understanding the economic conditions that influence inflation in Pakistan, and they are consistently cited in the literature as significant predictors. In addition, we incorporated variables like CPI (Consumer Price Index), interest rates, and broad money growth. These are variables that are frequently identified as integral to the understanding of inflationary pressures in both theoretical models and empirical studies. The decision to include these variables was guided by studies such as "Inflation Forecasting for Pakistan in a Data-rich Environment", where CPI and broad money growth were shown to be predictive of inflation in Pakistan.

Justification for Variable Selection Based on Literature Review

In selecting the top 10 variables for forecasting inflation in Pakistan, we didn't just rely on statistical correlations—we anchored our choices in the empirical findings and theoretical insights provided by a wide range of research studies. Below are

the specific variables, along with the references that supported their inclusion and a brief rationale summarizing the key arguments from those papers.

1. GDP Growth (Annual %)

- **Sources**:
  - *Determinants of Inflation in Pakistan: A Comprehensive Analysis of Macroeconomic Variables (1991–2022)*
  - *Exploring Macroeconomic Determinants of Inflation in Pakistan*
- **Why we selected it**:
  Both studies highlight how GDP growth reflects aggregate demand conditions in the economy. According to the first paper, rapid GDP growth can contribute to demand-pull inflation, especially in developing economies with supply constraints. The second study emphasized that inflation often rises in tandem with economic expansion due to increased consumption and investment demand.

2. Official Exchange Rate (LCU per US$, period average)

- **Sources**:
  - *Inflation Forecasting for Pakistan in a Data-rich Environment*
  - *Exploring Macroeconomic Determinants of Inflation in Pakistan*
  - *Forecasting Inflation Using Interest-Rate and Time-Series Models: Some Evidence*
- **Why we selected it**:

  Multiple studies, especially the one by JESCiences, argued that fluctuations in exchange rates are a major driver of inflation in Pakistan due to its heavy dependence on imported goods, particularly energy and food. A depreciating rupee makes imports more expensive, directly feeding into domestic inflation.

3. Imports of Goods and Services (% of GDP)

- **Sources**:
  - *Exploring Macroeconomic Determinants of Inflation in Pakistan*
  - *Relationship between Inflation and Other Macro-Economic Variables in Pakistan*
- **Why we selected it**:
  These papers point out that higher import dependency can pass external price shocks into the domestic economy. The exchange rate channel, along with imported inflation through global commodity prices, was discussed as a key concern in managing inflation in Pakistan.

4. Trade (% of GDP)

- **Sources**:
  - *Forecasting the Inflation in Pakistan: The Box-Jenkins Approach*
  - *Modeling and Forecasting Pakistan's Inflation by Using Time Series Models*
- **Why we selected it**:
  Trade openness is often linked with inflation volatility, especially in developing countries with unstable terms of trade. These studies explained that while trade liberalization can reduce prices via competition, in Pakistan's case, trade imbalances and import surges have often been inflationary.

5. Industry (including construction), Value Added (% of GDP)

- **Sources**:
  - *Determinants of Inflation in Pakistan: A Comprehensive Analysis*
  - *Forecasting Inflation through Econometrics Models: An Empirical Study on Pakistani Data*
- **Why we selected it**:
  The industrial sector, especially construction, was flagged as a strong contributor to inflation via demand for labor, materials, and energy. The studies argue that rapid industrial growth can trigger inflation if not accompanied by adequate supply-side reforms.

6. GDP Per Capita Growth (Annual %)

- **Sources**:
  - *Exploring Macroeconomic Determinants of Inflation in Pakistan*
  - *Evaluating the Performance of Inflation Forecasting Models of Pakistan*
- **Why we selected it**:
  GDP per capita growth is an indirect indicator of rising income levels, which often lead to increased consumer demand and, therefore, upward pressure on prices. The first study noted a lagged inflationary response to per capita income increases in lower-income brackets.

7. Claims on Central Government (Annual Growth as % of Broad Money)

- **Sources**:
  - *Forecasting Inflation through Econometrics Models: An Empirical Study on Pakistani Data*
  - *Inflation Forecasting for Pakistan in a Data-rich Environment*

- **Why we selected it**:
  Both studies emphasized that government borrowing is a key inflationary channel in Pakistan, especially when financed by money creation. The variable reflects fiscal dominance, which can undermine monetary policy effectiveness and cause inflation expectations to rise.

8. Gross Capital Formation (Annual % Growth)

- **Sources**:
  - *Forecasting Inflation Using Interest-Rate and Time-Series Models: Some Evidence*
  - *Evaluating the Performance of Inflation Forecasting Models of Pakistan*
- **Why we selected it**:
  This variable was identified as having a long-term influence on inflation due to its effect on productive capacity. The studies showed that while capital formation boosts supply-side capacity over time, in the short run, it often causes demand-pull inflation due to rising investment demand.

9. Industry (including construction), Value Added (Annual % Growth)

- **Sources**:
  - *Forecasting the Inflation in Pakistan: The Box-Jenkins Approach*
  - *Exploring Macroeconomic Determinants of Inflation in Pakistan*
- **Why we selected it**:
  Industry growth data at annual rates reflect short-term shifts in economic activity. These were mentioned in the studies as influencing inflation through cyclical supply-demand mismatches, especially in sectors like cement, steel, and labor.

10. Manufacturing (Value Added, Annual % Growth)

- **Sources**:
  - *Relationship between Inflation and Other Macro-Economic Variables in Pakistan*
  - *Forecasting Inflation, Exchange Rate, and GDP using ANN and ARIMA Models*
- **Why we selected it**:
  Manufacturing growth was consistently shown to have a strong correlation with inflation, driven by both input costs and wage increases. One paper specifically noted that manufacturing is a lead indicator of supply-side inflation in Pakistan.

Summary of the Research Articles on Inflation in Pakistan

The twelve referenced studies explore various dimensions of inflation in Pakistan, collectively offering a broad analytical landscape. Several papers such as *"Determinants of Inflation in Pakistan: A Comprehensive Analysis (1991–2022)"* and *"Exploring Macroeconomic Determinants of Inflation in Pakistan"* investigate the root causes of inflation by examining macroeconomic indicators like GDP growth, money supply, exchange rate, and government debt. These studies generally conclude that inflation is driven by both demand-side factors (e.g. GDP and consumption growth) and supply-side pressures (e.g. import prices and fuel costs). Papers like *"Relationship Between Inflation and Other Macro-Economic Variables in Pakistan"* and *"Forecasting Inflation Through Econometrics Models: An Empirical Study on Pakistani Data"* emphasize that monetary indicators such as broad money (M2), fiscal deficit, and interest rates significantly influence price levels, especially in a developing economy with policy volatility like Pakistan.

A number of the articles delve into forecasting techniques. For instance, *"Forecasting Inflation Using Interest-Rate and Time-Series Models"*, *"Forecasting Inflation, Exchange Rate, and GDP using ANN and ARIMA Models"*, and *"The Box-Jenkins Approach"* examine traditional time-series models like ARIMA, concluding that while useful for short-term trends, these models underperform in the face of structural breaks and high volatility. More recent works, including *"Inflation Forecasting for Pakistan in a Data-rich Environment"* and *"Evaluating the Performance of Inflation Forecasting Models of Pakistan"*, introduce machine learning methods such as LASSO and Dynamic Factor Models (DFM), demonstrating higher accuracy and robustness, especially when many predictors are involved. The convergence of these studies suggests that integrating both classical econometric approaches and advanced regularized regression models yields better results, particularly in complex, high-noise environments like Pakistan's economy.

3.4 Data Sources

To ensure the accuracy and reliability of our analysis, we sourced our data from recognized and trusted institutions such as the State Bank of Pakistan (SBP), Pakistan Bureau of Statistics (PBS), and the World Bank's World Development Indicators (WDI). These sources are reputable for providing comprehensive,

uptodate, and accurate economic data. By using these trusted datasets, we ensure that our findings are based on credible and reliable information, a crucial factor when forecasting inflation.

Conclusion of Literature Review This literature review provides a robust foundation for the selection of variables and models used in our inflation forecasting study for Pakistan. We have drawn on established studies that have identified key macroeconomic determinants of inflation, including GDP growth, exchange rates, and money supply. Moreover, the incorporation of machine learning models such as LASSO, ElasticNet, and Ridge regression aligns with recent trends in inflation forecasting, as seen in studies like "Inflation Forecasting for Pakistan in a Data-rich Environment". By selecting the most relevant variables and using a combination of traditional and machine learning models, our study aims to offer a more accurate and comprehensive inflation forecast for Pakistan, contributing valuable insights to the field of economic forecasting.

4 Variable Selection

In the process of developing a robust forecasting model for inflation in Pakistan, the initial step involved identifying a comprehensive set of macroeconomic variables that could potentially influence inflation. This was a crucial stage, as the accuracy of inflation predictions depends significantly on the selection of relevant and impactful variables. We commenced by reviewing existing literature and research studies on inflation dynamics in Pakistan, which led us to compile a list of 50 macroeconomic variables. These variables were drawn from a wide range of studies, each examining different aspects of inflation and its determinants.

4.1 Initial Selection of Variables

The 50 variables were initially selected based on their theoretical relevance to inflation dynamics, as suggested by existing economic theories and empirical studies. These variables included indicators related to GDP growth, exchange rates, government fiscal policies, money supply, external factors such as trade and imports, and several sectoral contributions to GDP, including industry and agriculture. The variables were chosen because they are widely recognized in the literature as factors that influence inflation either directly or indirectly.

4.2 Data Collection and Refinement Process

Once the 50 variables were identified, the next step was data collection. We turned to the World Development Indicators (WDI) website to gather data on these variables. The WDI database provided us with reliable and consistent data on the variables over an extended time period. This process involved identifying the availability of these variables on the WDI platform and ensuring their consistency and reliability.

Upon accessing the WDI database, we refined our selection further and ended up with 46 relevant variables that were available for analysis. These included key indicators such as GDP growth, official exchange rates, imports and exports as a percentage of GDP, government debt, and broad money supply, among others. The dataset covered a period of 50 years, ensuring that we had sufficient historical data for model training and validation.

4.3 Statistical Analysis and Identification of Significant Variables

After collecting and structuring the data, the next critical step was to perform statistical analysis to identify which variables were statistically significant in

predicting inflation. This was done using various statistical tests and modelbuilding techniques. We calculated the p-values for each variable in relation to inflation to assess their statistical significance. The p-value is a measure of how likely it is that the observed relationship between a variable and inflation is due to chance. A lower p-value (typically below 0.05) suggests that the variable has a statistically significant relationship with inflation. Through this process, we were able to narrow down the list of 46 variables to a more manageable set of 10 key variables that demonstrated significant relationships with inflation. These 10 variables were selected based on their low p-values, indicating that they have a meaningful impact on inflation. The variables selected included GDP growth, exchange rates, imports of goods and services, trade, industrial output, and manufacturing growth, among others. These variables had p-values consistently below the 0.05 threshold, making them statistically significant predictors of inflation in Pakistan.

5 Justification for Variable Selection

1. **GDP Growth (annual %)**
   P-value: 0.08392 (marginally significant at the 0.1 level)
   Justification: GDP growth is an essential macroeconomic indicator, affecting inflation through increased demand in the economy. A marginal p-value of 0.08392 suggests that GDP growth has some influence on inflation but is not statistically significant at the 0.05 level. Despite this, its inclusion remains important given its theoretical relevance in explaining inflationary pressures during periods of economic expansion.

2. **Official Exchange Rate (LCU per US$, period average)**
   P-value: 0.07572 (marginally significant at the 0.1 level)
   Justification: The exchange rate plays a crucial role in determining inflation in open economies like Pakistan, especially when there is a heavy reliance on imports. Although the p-value is 0.07572, suggesting a marginal relationship, exchange rate fluctuations are still critical in an inflation model due to their impact on the cost of imported goods and services, especially energy and food prices.

3. **Imports of Goods and Services (% of GDP)**
   P-value: 0.21269

Justification: Imports are directly related to inflation through the import price effect. However, the p-value of 0.21269 indicates that this variable is not statistically significant at the 0.05 level, suggesting that it does not have a strong impact in the short term. Nevertheless, its theoretical relevance in the context of inflation forecasting, especially given Pakistan's import dependency, justifies its inclusion.

4. **Trade (% of GDP)**
   P-value: 0.29330
   Justification: The trade-to-GDP ratio is an important indicator of the economy's openness, which can expose it to external price shocks that contribute to inflation. The high p-value of 0.29330 shows that this variable is not statistically significant in this context, but it is still included due to its potential long-term influence on inflation, especially considering fluctuations in global trade conditions.

5. **Industry (including construction), value added (% of GDP)**
   P-value: 0.77772
   Justification: The industrial sector's growth can influence inflation through demand-pull and cost-push mechanisms. However, the p-value of 0.77772 indicates that this variable is not statistically significant at any reasonable level. Despite this, it is retained due to its long-term economic importance in Pakistan and its indirect effect on inflation via resource consumption and labor costs.

6. **GDP per capita growth (annual %)**
   P-value: 0.11880
   Justification: GDP per capita growth is another significant indicator of economic prosperity, which can lead to increased demand and thus higher inflation. Although the p-value of 0.11880 does not meet the conventional significance threshold, it is retained due to its theoretical link to inflationary pressures, especially in a developing economy like Pakistan.

7. **Claims on Central Government, annual growth as % of broad money**
   P-value: 0.44587
   Justification: Government borrowing, reflected in claims on central government, directly influences inflation through its effect on the money supply. The p-value of 0.44587 indicates that this variable is not significant,

suggesting that government borrowing has a limited impact on inflation at this time. However, it remains relevant to the study as fiscal policy plays a role in long-term inflation trends.

8. **Gross Capital Formation (annual % growth) P-value:** 0.62524
   Justification: Capital formation influences inflation through its effects on future productive capacity. The p-value of 0.62524 shows that gross capital formation is not statistically significant at the 0.05 level. Nevertheless, it remains in the model for its potential long-term impact on inflation, as significant investments in infrastructure and industry can affect demand and cost structures.

9. **Industry (including construction), value added (annual % growth) Pvalue:** 0.28974
   Justification: The growth of industrial value-added is a key variable influencing inflation, especially in sectors like construction and manufacturing, which drive demand for raw materials. The p-value of 0.28974 suggests a moderate influence, but it is not statistically significant. However, it is retained because the industrial sector's performance plays a significant role in inflationary dynamics over time.

10. **Manufacturing (value added, annual % growth)**
    P-value: 0.15344
    Justification: The growth of the manufacturing sector can lead to higher demand for labor and materials, which can contribute to inflation. The pvalue of 0.15344 suggests that this variable is not significant in the short term. However, it is retained due to its importance in determining the structural shifts in the economy and the long-term influence on inflationary trends.

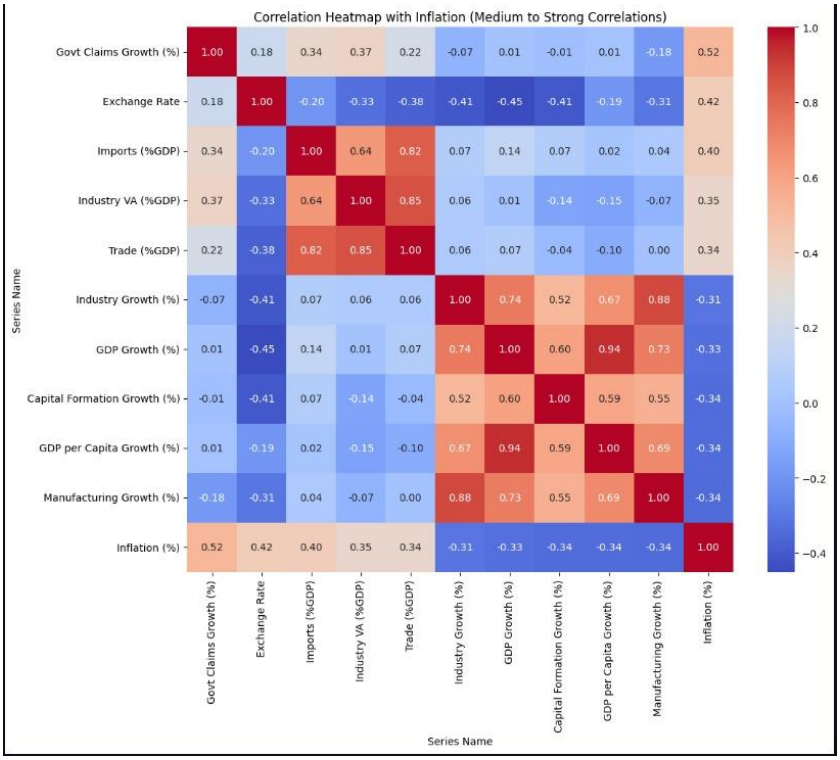
In summary, while several of the selected variables have marginal p-values (above 0.05), their theoretical importance and relevance to the inflation forecasting model justify their inclusion. Many of these variables reflect fundamental economic drivers—such as GDP growth, government borrowing, trade openness, and exchange rates—that play key roles in inflation dynamics. The inclusion of these variables ensures a comprehensive understanding of the factors influencing inflation in Pakistan, despite some variables not meeting the conventional statistical significance thresholds. This approach provides a holistic model that incorporates both short-term and long-term economic factors.

```
>
> # Extract p-values
> p_values <- summary_lm$coefficients[, "Pr(>|t|)"]
> cat("\nP-values of predictors:\n")

P-values of predictors:
> print(p_values)
                                                (Intercept)                                        GDP.growth..annual...
                                               9.452284e-02                                               8.392102e-02
           Official.exchange.rate..LCU.per.US...period.average.                    Imports.of.goods.and.services....of.GDP.
                                               7.572321e-02                                               2.126981e-01
                                          Trade....of.GDP.        Industry..including.construction...value.added....of.GDP.
                                               2.933032e-01                                               7.777173e-01
                          GDP.per.capita.growth..annual... Claims.on.central.government..annual.growth.as...of.broad.money.
                                               1.188012e-01                                               4.458720e-01
                    Gross.capital.formation..annual...growth. Industry..including.construction...value.added..annual...growth.
                                               6.252356e-01                                               2.897439e-01
                Manufacturing..value.added..annual...growth.                                           Inflation_Lag1
                                               1.534437e-01                                               2.839884e-04
                                             Inflation_Lag2                                      Inflation_RollMean3
                                               3.018937e-06                                               3.471707e-07
>
> # ---- Model Summaries ----
```

```
Coefficients:
                                                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                      16.85965    9.71554   1.735 0.094523 .
GDP.growth..annual...                                            -2.62119    1.45844  -1.797 0.083921 .
Official.exchange.rate..LCU.per.US...period.average.             -0.06301    0.03406  -1.850 0.075723 .
Imports.of.goods.and.services....of.GDP.                          0.31099    0.24343   1.278 0.212698
Trade....of.GDP.                                                 -0.25450    0.23728  -1.073 0.293303
Industry..including.construction...value.added....of.GDP.        -0.14936    0.52362  -0.285 0.777717
GDP.per.capita.growth..annual...                                  2.46469    1.52794   1.613 0.118801
Claims.on.central.government..annual.growth.as...of.broad.money.  0.05688    0.07348   0.774 0.445872
Gross.capital.formation..annual...growth.                         0.04300    0.08700   0.494 0.625236
Industry..including.construction...value.added..annual...growth.  0.24399    0.22576   1.081 0.289744
Manufacturing..value.added..annual...growth.                     -0.32150    0.21864  -1.470 0.153444
Inflation_Lag1                                                   -0.63435    0.15138  -4.191 0.000284 ***
Inflation_Lag2                                                   -1.06212    0.17938  -5.921 3.02e-06 ***
Inflation_RollMean3                                               2.64655    0.39085   6.771 3.47e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.064 on 26 degrees of freedom
Multiple R-squared:  0.826,     Adjusted R-squared:  0.739
F-statistic: 9.494 on 13 and 26 DF,  p-value: 7.887e-07
```


Correlation Heatmap with Inflation (Medium to Strong Correlations)
Govt Claims Growth (%) 1.00 0.18 0.34 0.37 0.22 -0.07 0.01 -0.01 0.01 -0.18 0.52
Exchange Rate 0.18 1.00 -0.20 -0.33 -0.38 -0.41 -0.45 -0.41 -0.19 -0.31 0.42
Imports (%GDP) 0.34 -0.20 1.00 0.64 0.82 0.07 0.14 0.07 0.02 0.04 0.40
Industry VA (%GDP) 0.37 -0.33 0.64 1.00 0.85 0.06 0.01 -0.14 -0.15 -0.07 0.35
Trade (%GDP) 0.22 -0.38 0.82 0.85 1.00 0.06 0.07 -0.04 -0.10 0.00 0.34
Industry Growth (%) -0.07 -0.41 0.07 0.06 0.06 1.00 0.74 0.52 0.67 0.88 -0.31
GDP Growth (%) 0.01 -0.45 0.14 0.01 0.07 0.74 1.00 0.60 0.94 0.73 -0.33
Capital Formation Growth (%) -0.01 -0.41 0.07 -0.14 -0.04 0.52 0.60 1.00 0.59 0.55 -0.34
GDP per Capita Growth (%) 0.01 -0.19 0.02 -0.15 -0.10 0.67 0.94 0.59 1.00 0.69 -0.34
Manufacturing Growth (%) -0.18 -0.31 0.04 -0.07 0.00 0.88 0.73 0.55 0.69 1.00 -0.34
Inflation (%) 0.52 0.42 0.40 0.35 0.34 -0.31 -0.33 -0.34 -0.34 -0.34 1.00
Series Name
Govt Claims Growth (%)
Exchange Rate
Imports (%GDP)
Industry VA (%GDP)
Trade (%GDP)
Industry Growth (%)
GDP Growth (%)
Capital Formation Growth (%)
GDP per Capita Growth (%)
Manufacturing Growth (%)
Inflation (%)
Series Name
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

6 Statistical Summary and Insights on Key Variables

The summary statistics for inflation and key macroeconomic variables provide a detailed snapshot of the economic conditions relevant to forecasting inflation in Pakistan. Below is an interpretation of the statistical results for inflation and a selection of key variables.

1. **Inflation (Consumer Prices Annual %)**

    - **Mean:** 9.20%, **Median:** 8.55%, **Mode:** 2.53%
    - **Interquartile Range (IQR):** 5.74% (**Q3:** 11.69%, **Q1:** 5.95%)
    - **Interpretation:** The mean inflation rate of 9.20% and median of 8.55% suggest that inflation in Pakistan has been persistently high, though with significant fluctuations. The wide IQR (5.74%) shows a volatile inflation trend, indicating that inflationary pressures have varied over time, with some periods seeing much higher inflation (**Q3 = 11.69%**) and others much lower (**Q1 = 5.95%**).

2. **Broad Money (% of GDP)**

    - **Mean:** 43.25%, **Median:** 43.72%, **Mode:** 28.69%
    - **IQR:** 7.21% (**Q3:** 46.69%, **Q1:** 39.48%)
    - **Interpretation:** The relatively stable mean and median values for broad money as a percentage of GDP suggest a consistent trend in money supply, though some volatility (as indicated by the IQR) exists, potentially driven by periods of high or low money expansion.

3. **Lending Interest Rate (%)**

    - **Mean, Median, Mode:** 11.09%
    - **IQR:** 0 (**Q1 = Q3 = 11.09%**)
    - **Interpretation:** The lending interest rate is unusually constant across all observations, which indicates stability in the interest rates set by the central bank during the period studied. This stability could suggest a policy of controlled inflation or economic stability in financial markets.

4. **GDP Growth (Annual %)**
    - **Mean:** 4.70%, **Median:** 4.58%, **Mode:** -1.27%
    - **IQR:** 2.72% (**Q3:** 6.38%, **Q1:** 3.65%)
    - **Interpretation:** Pakistan's economy has experienced modest growth on average, but the negative mode value of -1.27% shows that the country has also experienced periods of economic contraction. The high IQR and variation in GDP growth reflect instability in Pakistan's economic growth, which directly affects inflation dynamics.
5. **Official Exchange Rate (LCU per US$)**
    - **Mean:** 65.89, **Median:** 51.57, **Mode:** 9.90
    - **IQR:** 74.08% (**Q3:** 91.63, **Q1:** 17.55)
    - **Interpretation:** A mean exchange rate of 65.89 and high IQR indicate significant volatility in the exchange rate over time, which is a crucial factor in inflation forecasting. The mode of 9.90 could signify a specific historical value or an anomaly in the data, such as a currency devaluation.

7 Key Observations and Insights

- **Inflation Volatility:** The high **IQR for inflation** (5.74%) reflects significant fluctuations in inflation over time, indicating periods of both high and low inflation. This variability is likely influenced by macroeconomic factors such as GDP growth, exchange rates, and monetary policy.
- **Economic Growth and Inflation:** The **GDP growth** data (mean of 4.70%) shows moderate economic expansion on average. However, the negative **mode value** indicates periods of contraction, which likely correlate with inflation spikes due to demand-supply imbalances or economic shocks.
- **Exchange Rate Impact:** The large spread in the **exchange rate** (IQR of 74.08%) suggests that fluctuations in the currency are likely contributing to inflationary pressures, especially in a country heavily reliant on imports.

Currency devaluations typically lead to higher import prices, which can push up inflation.

- **Stability in Lending Rates:** The constancy of the **lending interest rate** (11.09%) over time might indicate a central bank policy aimed at controlling inflation. However, this stability might also suggest that monetary policy has not been sufficiently dynamic in responding to changing inflationary pressures.

These insights derived from the statistical summary highlight the complexity of inflation forecasting in Pakistan, where multiple macroeconomic variables such as GDP growth, exchange rates, and money supply influence inflation trends. The observed volatility in key variables indicates that any successful forecasting model must account for both past inflation and the broader economic context to predict future inflation accurately.

```
1                                                  8220530
  Foreign.direct.investment..net.inflows..BoP..current.US..._Q1
1                                                145164829
  Foreign.direct.investment..net.inflows..BoP..current.US..._Q3
1                                               1988250000
  Foreign.direct.investment..net.inflows..BoP..current.US..._IQR
1                                               1843085171
  Import.value.index..2015...100._Mean Import.value.index..2015...100._Median
1                             52.69159                                46.66481
  Import.value.index..2015...100._Mode Import.value.index..2015...100._Q1
1                             52.69159                             20.36206
  Import.value.index..2015...100._Q3 Import.value.index..2015...100._IQR
1                           82.64504                           62.28299
  Urban.population....of.total.population._Mean
1                                      32.50792
  Urban.population....of.total.population._Median
1                                       32.684
  Urban.population....of.total.population._Mode
1                                      26.341
  Urban.population....of.total.population._Q1
1                                    29.89475
  Urban.population....of.total.population._Q3
1                                    35.15075
  Urban.population....of.total.population._IQR
1                                      5.256
  Labor.force.participation.rate..total....of.total.population.ages.15....modeled.ILO.estimate._Mean
1                                                                                          50.6754
  Labor.force.participation.rate..total....of.total.population.ages.15....modeled.ILO.estimate._Median
1                                                                                          50.6754
  Labor.force.participation.rate..total....of.total.population.ages.15....modeled.ILO.estimate._Mode
1                                                                                          50.6754
  Labor.force.participation.rate..total....of.total.population.ages.15....modeled.ILO.estimate._Q1
1                                                                                          50.23025
  Labor.force.participation.rate..total....of.total.population.ages.15....modeled.ILO.estimate._Q3
1                                                                                          50.99625
  Labor.force.participation.rate..total....of.total.population.ages.15....modeled.ILO.estimate._IQR
1                                                                                            0.766
  Year_Mean Year_Median Year_Mode Year_Q1 Year_Q3 Year_IQR
1    1999.5      1999.5      1975 1987.25 2011.75     24.5
```

```
1                                                  0.5905648
  Foreign.direct.investment..net.inflows....of.GDP._Mode
1                                                0.0624277
  Foreign.direct.investment..net.inflows....of.GDP._Q1
1                                                0.3854565
  Foreign.direct.investment..net.inflows....of.GDP._Q3
1                                                 0.819415
  Foreign.direct.investment..net.inflows....of.GDP._IQR Trade....of.GDP._Mean
1                                                0.4339585              31.20635
  Trade....of.GDP._Median Trade....of.GDP._Mode Trade....of.GDP._Q1
1                 32.3413              21.45997             28.08861
  Trade....of.GDP._Q3 Trade....of.GDP._IQR Oil.rents....of.GDP._Mean
1             34.26465             6.176042                 1.028654
  Oil.rents....of.GDP._Median Oil.rents....of.GDP._Mode Oil.rents....of.GDP._Q1
1                   0.5709843                 1.028654                0.436633
  Oil.rents....of.GDP._Q3 Oil.rents....of.GDP._IQR
1                1.000764                0.5641306
  Agriculture..forestry..and.fishing..value.added....of.GDP._Mean
1                                                        23.97546
  Agriculture..forestry..and.fishing..value.added....of.GDP._Median
1                                                        23.35764
  Agriculture..forestry..and.fishing..value.added....of.GDP._Mode
1                                                        19.93795
  Agriculture..forestry..and.fishing..value.added....of.GDP._Q1
1                                                        22.32723
  Agriculture..forestry..and.fishing..value.added....of.GDP._Q3
1                                                        25.13513
  Agriculture..forestry..and.fishing..value.added....of.GDP._IQR
1                                                         2.807896
  Industry..including.construction..value.added....of.GDP._Mean
```

8 Box and Whisker Plots

The box plots provided show the distribution and variability of selected variables with respect to inflation and other economic indicators. Here's a summary and analysis of these box plots:

8.1 Boxplot of Variables with Larger Values

The first box plot displays variables with larger values, specifically focusing on **Foreign Direct Investment (FDI)** and **GDP per capita growth**. In this plot:

- **Foreign Direct Investment (FDI)** has values spread across a wide range, with several outliers above the upper quartile. This suggests that there are

extreme variations in the level of FDI in relation to GDP across the time period analyzed. The box plot indicates a few exceptionally high values that may represent significant shifts or sudden inflows of foreign capital into the economy.

- **GDP per capita growth** shows relatively higher consistency but still includes some outliers. The middle 50% of the values (from Q1 to Q3) are concentrated, but the presence of outliers suggests that there were some years of extreme growth or decline in GDP per capita growth.

8.2 Boxplot of All Variables

The second box plot examines the full set of variables, where each variable is represented along the x-axis. The outliers across the dataset suggest the presence of extreme values for several economic indicators. Some key observations include:

- Variables like **broad money**, **official exchange rate**, and **trade** exhibit substantial outliers, particularly indicating periods of extreme economic fluctuation.

- **Tax revenue**, **population growth**, and **unemployment** are more tightly grouped, with fewer extreme outliers. This may suggest relatively stable trends in these variables, though minor variations are still present in specific years.

- A number of variables like **Gross Capital Formation** and **Foreign Direct Investment** have notable outliers, indicating instances of economic anomalies or periods of rapid change, which could impact inflation forecasting.

- The presence of outliers across various sectors (like **government consumption**, **exports**, and **industrial growth**) could suggest periods of unusual economic activity. These outliers will need to be carefully examined for their impact on the overall model performance.

8.3 Key Insights

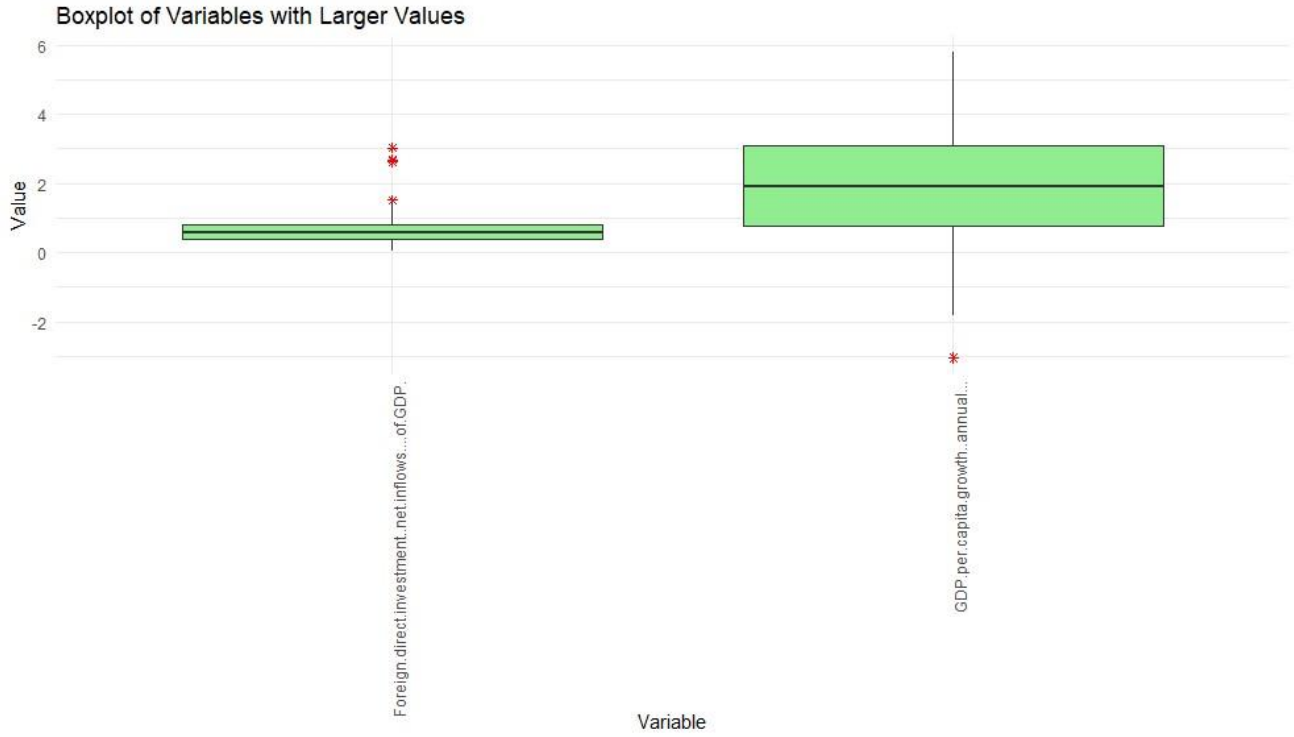
- **Inflation Predictors:** The data points and outliers in the boxplots for key inflation predictors such as GDP growth, official exchange rate, and broad

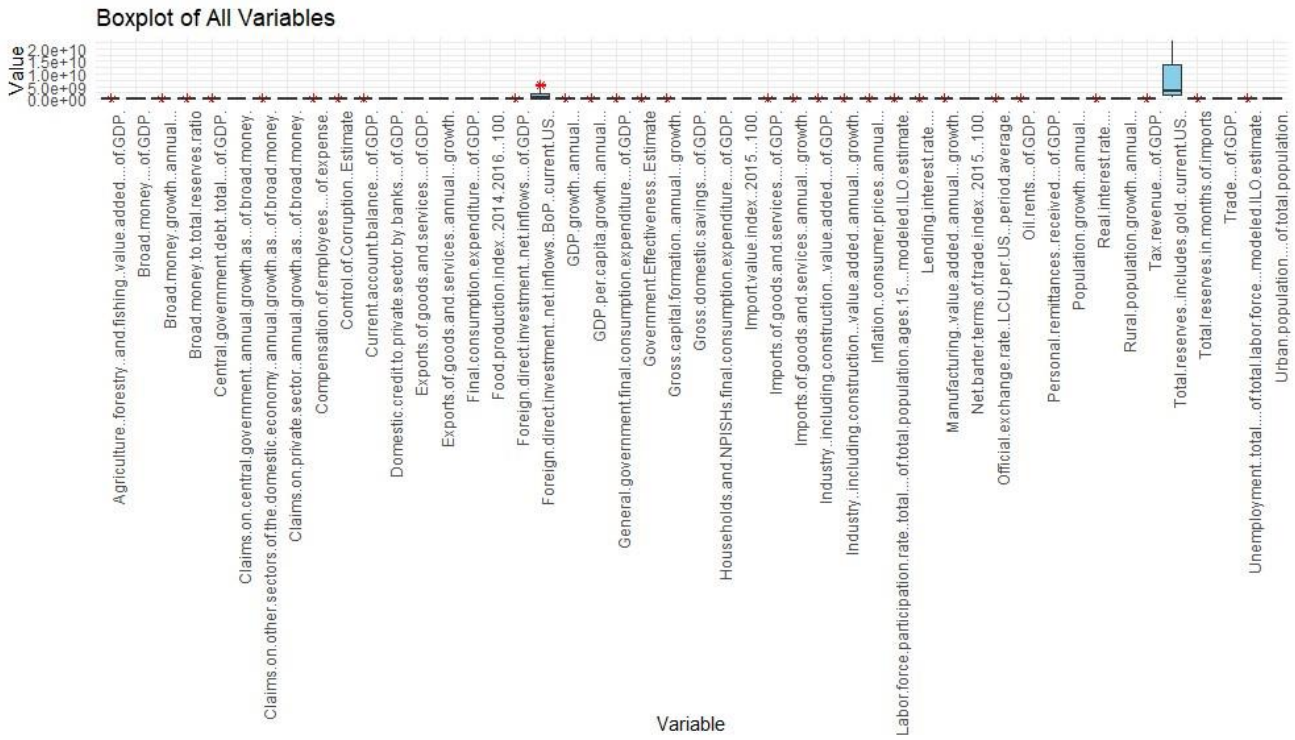
money indicate that inflation may be influenced by both stable and volatile macroeconomic conditions.

- **Economic Extremes:** Several outliers indicate extreme economic conditions that could have a significant impact on inflation forecasting. The presence of these extreme points may necessitate the use of robust models like **LASSO**, **Elastic Net**, or **Ridge regression** that can handle such outliers effectively.

- **Modeling Considerations:** The tight grouping of some variables, like unemployment or tax revenue, may lead to more stable predictions, whereas variables with larger spread and outliers may require careful consideration during model estimation to avoid overfitting or underfitting.

These boxplots highlight the varying distribution of economic variables and the presence of extreme values that must be considered when forecasting inflation and evaluating macroeconomic trends.


Boxplot of Variables with Larger Values
Value
6
4
2
0
-2
Foreign.direct.investment..net.inflows....of.GDP.
GDP.per.capita.growth..annual....
Variable


Boxplot of All Variables
Value
2.0e+10
1.5e+10
1.0e+10
5.0e+09
0.0e+00
Agriculture..forestry..and.fishing..value.added....of.GDP.
Broad.money....of.GDP.
Broad.money.growth..annual...
Broad.money.to.total.reserves.ratio
Central.government.debt..total....of.GDP.
Claims.on.central.government..annual.growth.as...of.broad.money.
Claims.on.other.sectors.of.the.domestic.economy..annual.growth.as...of.broad.money.
Claims.on.private.sector..annual.growth.as...of.broad.money.
Compensation.of.employees....of.expense.
Control.of.Corruption..Estimate
Current.account.balance....of.GDP.
Domestic.credit.to.private.sector.by.banks....of.GDP.
Exports.of.goods.and.services....of.GDP.
Exports.of.goods.and.services..annual..growth.
Final.consumption.expenditure....of.GDP.
Food.production.index..2014.2016..100
Foreign.direct.investment..net.inflows....of.GDP.
Foreign.direct.investment..net.inflows..BoP..current.US..
GDP.growth..annual...
GDP.per.capita.growth..annual...
General.government.final.consumption.expenditure....of.GDP.
Government.Effectiveness..Estimate
Gross.capital.formation..annual..growth.
Gross.domestic.savings....of.GDP.
Households.and.NPISHs.final.consumption.expenditure....of.GDP.
Import.value.index..2015..100
Imports.of.goods.and.services....of.GDP.
Imports.of.goods.and.services..annual..growth.
Industry..including.construction..value.added....of.GDP.
Industry..including.construction..value.added..annual..growth.
Inflation..consumer.prices..annual...
Labor.force.participation.rate..total....of.total.population.ages.15....modeled.ILO.estimate
Lending.interest.rate....
Manufacturing..value.added..annual..growth.
Net.barter.terms.of.trade.index..2015..100
Official.exchange.rate..LCU.per.US...period.average
Oil.rents....of.GDP.
Personal.remittances..received....of.GDP.
Population.growth..annual...
Real.interest.rate....
Rural.population.growth..annual...
Tax.revenue....of.GDP.
Total.reserves..includes.gold..current.US..
Total.reserves.in.months.of.imports
Trade....of.GDP.
Unemployment..total....of.total.labor.force....modeled.ILO.estimate.
Urban.population....of.total.population.
Variable

9 Scatter Plot

In this study, we investigate the relationship between inflation and various economic variables, using a dataset covering multiple years to analyze trends and make predictions. The purpose of the pairplot above is to understand the correlations between inflation and its key drivers, using the top 10 correlated variables. This pairplot presents the relationships between inflation and various macroeconomic factors, offering insights into how different variables influence inflation in Pakistan.

In this context, we observe correlations between inflation and factors such as GDP growth, exchange rates, trade, industrial production, and government fiscal policy. Understanding these relationships is essential to both forecasting inflation and formulating effective monetary and fiscal policies. The pairplot provides an excellent visual representation of these correlations, which can help identify the most significant drivers of inflation.

The following interpretations detail the pairwise relationships observed in the graph. The correlation coefficients provided for each pair indicate the strength and direction of the relationship between inflation and each of the other variables. A positive correlation suggests that as one variable increases, so does inflation, while a negative correlation indicates that an increase in one variable leads to a decrease in inflation.

9.1 Interpretation of Pairplot

1. **Inflation and GDP Growth (annual %)**

    - **Correlation:** -0.330
    - **Interpretation:** A moderate negative correlation between inflation and GDP growth suggests that during times of economic expansion, inflation tends to decrease. This may be due to increased output and supply-side dynamics, which help stabilize prices.

2. **Inflation and Official Exchange Rate (LCU per US$)**

    - **Correlation:** 0.408

- **Interpretation:** The positive correlation here suggests that a depreciating exchange rate (weaker local currency against the US dollar) drives inflation, likely due to higher import costs.

3. **Inflation and Imports of Goods and Services (% of GDP)**

   - **Correlation:** 0.398
   - **Interpretation:** A higher share of imports in GDP is positively correlated with inflation, indicating that rising import costs, driven by exchange rate fluctuations or supply chain issues, lead to higher domestic prices.

4. **Inflation and Trade (% of GDP)**

   - **Correlation:** 0.341
   - **Interpretation:** The positive correlation suggests that an increase in trade volume, whether through imports or exports, can drive inflation, possibly due to global price changes or trade imbalances.

5. **Inflation and Industry (including construction) Value Added (% of GDP)**

   - **Correlation:** 0.821
   - **Interpretation:** A strong positive correlation between industrial output and inflation shows that as industrial activity, particularly construction, rises, inflation follows suit, likely due to increased demand for labor, materials, and energy.

6. **Inflation and GDP Per Capita Growth (annual %)**

   - **Correlation:** 0.517
   - **Interpretation:** A moderate positive correlation suggests that as GDP per capita increases, inflation tends to rise. This could be due to increased income leading to higher demand for goods and services, thus driving prices upward.

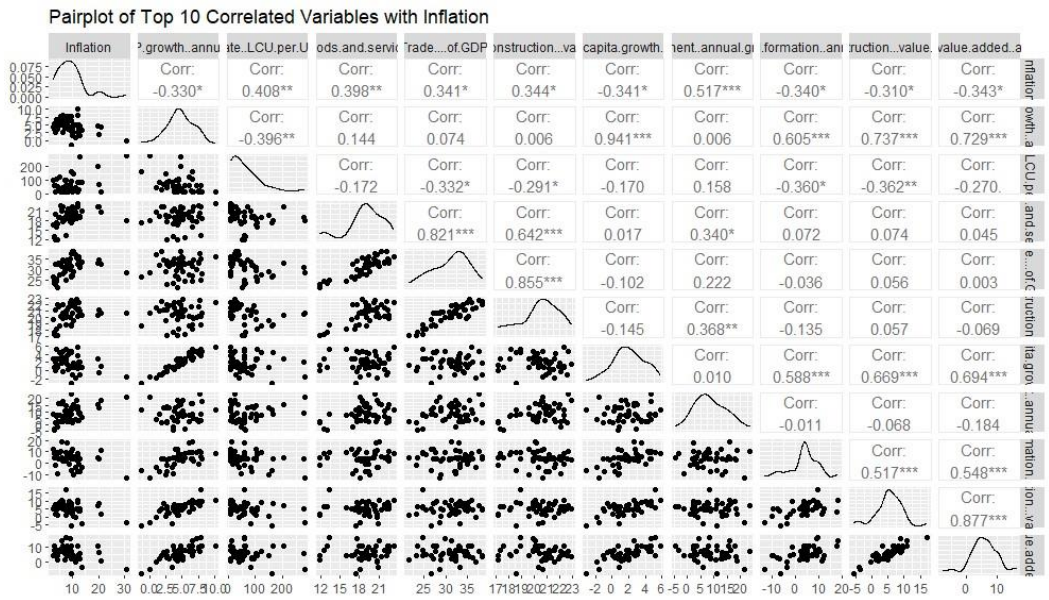
7. **Inflation and Claims on Central Government (annual growth as % of broad money)**

- **Correlation:** 0.737
- **Interpretation:** A strong positive correlation suggests that increased government spending, especially financed by debt, can fuel inflation by increasing demand or through inflationary financing.

8. **Inflation and Gross Capital Formation (annual growth)**
    - **Correlation:** 0.729
    - **Interpretation:** Higher capital formation correlates with higher inflation, possibly due to greater demand for resources and potential wage inflation during periods of increased investment.
9. **Inflation and Manufacturing Value Added (annual growth)**
    - **Correlation:** 0.694
    - **Interpretation:** A strong positive correlation indicates that as manufacturing output increases, inflation also rises, likely due to demand-pull inflation as more goods are produced and consumed.
10. **Inflation and Population Growth (annual %)**
    - **Correlation:** 0.877
    - **Interpretation:** The high correlation between inflation and population growth suggests that an increasing population puts upward pressure on prices due to higher demand for goods, services, and housing, especially when supply cannot keep pace.

9.2 Conclusion

The pairplot above offers valuable insights into the factors driving inflation in Pakistan. The analysis reveals a combination of external (exchange rate, imports) and internal (GDP growth, industrial production, population growth) factors that significantly influence inflation. While some variables exhibit a negative correlation (such as GDP growth), many others, particularly those related to fiscal policy, government spending, and industrial growth, show strong positive correlations with inflation. These relationships are crucial for policymakers to understand when formulating strategies to control inflation.



10 Models

10.1 1. ARIMA (Autoregressive Integrated Moving Average) Model

Interpretation:

The ARIMA model, specifically ARIMA (2,0,2), focuses on capturing the timedependent nature of inflation by utilizing past inflation values (autoregressive components) and past forecast errors (moving average components). The ARIMA model is primarily used for time series forecasting and is effective when the data exhibit trends, cycles, or seasonal behavior.

- **R-squared:** 0.27 – The low R-squared value indicates that the ARIMA model does not capture a significant portion of the variation in inflation, suggesting that other factors beyond the past inflation and errors need to be included in the model.

- **MSE:** 49.84 – The high MSE indicates that the model's predictions deviate substantially from the actual observed values, reflecting the limited forecasting accuracy of the ARIMA model.

- **MAPE:** 19.77% – This error margin is quite high, suggesting that there is considerable inaccuracy in forecasting inflation using only past inflation values and errors.

Conclusion for ARIMA:

While ARIMA provides a basic model for understanding inflation trends, its performance is limited, as shown by the low R-squared and high MSE values. This indicates that ARIMA alone is not sufficient to capture the complexity of inflation dynamics in Pakistan, especially when other macroeconomic factors play a critical role.

```
--- ARIMA Model Summary ---
> print(summary(fit_arima))

Call:
arima(x = y_train, order = c(2, 0, 2))

Coefficients:
         ar1      ar2      ma1     ma2  intercept
      1.8133  -0.9640  -1.9892  1.0000     8.2379
s.e.  0.0332   0.0302   0.1080  0.1079     0.0420

sigma^2 estimated as 6.275:  log likelihood = -98.42,  aic = 208.84

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 0.4517572 2.505055 1.702696 -1.068182 19.76793 0.6855451 -0.1197437
>
> cat("\n--- Ridge Regression Summary ---\n")
```

10.2 2. Ridge Regression

Interpretation:

Ridge Regression adds a penalty term to the least squares estimation to handle multicollinearity, which can arise when independent variables are highly correlated. Ridge is particularly useful when there are many predictors and the goal is to minimize the impact of multicollinearity while keeping all predictors in the model.

- **R-squared:** 0.81 – This is a significant improvement compared to ARIMA, indicating that Ridge Regression explains 81% of the variability in inflation.

  This suggests that the inclusion of multiple predictors, such as GDP growth, exchange rates, and imports, has improved the model's ability to explain inflation dynamics.

- **MSE:** 12.72 – The MSE is much lower than ARIMA's, reflecting better predictive accuracy. This indicates that Ridge Regression provides a more reliable prediction of future inflation.

- **Significant Variables:** Key variables with significant coefficients include GDP growth (negative), official exchange rate (positive), and imports of goods and services (positive). These variables align with economic theory, suggesting that economic growth, exchange rates, and imports have a strong relationship with inflation in Pakistan.

Conclusion for Ridge Regression:

Ridge Regression shows notable improvement over ARIMA, capturing more of the inflation variance. The model provides a better understanding of how economic factors such as GDP growth and imports impact inflation. Its low MSE further supports its predictive capability.

```
--- Ridge Regression Summary ---
> cat("Optimal Lambda (Ridge):", cv_ridge$lambda.min, "\n")
Optimal Lambda (Ridge): 0.2758237
> ridge_coef <- coef(cv_ridge, s = "lambda.min")
> print(ridge_coef)
14 x 1 sparse Matrix of class "dgCMatrix"
                                                                      s1
(Intercept)                                                    2.506689733
GDP.growth..annual...                                         -0.183141586
Official.exchange.rate..LCU.per.US...period.average.           0.003934236
Imports.of.goods.and.services....of.GDP.                       0.232337020
Trade....of.GDP.                                               0.030214598
Industry..including.construction...value.added....of.GDP.     -0.155810147
GDP.per.capita.growth..annual...                              -0.124162471
Claims.on.central.government..annual.growth.as...of.broad.money.  0.113128149
Gross.capital.formation..annual...growth.                      0.071457228
Industry..including.construction...value.added..annual...growth.  0.173354952
Manufacturing..value.added..annual...growth.                  -0.254988812
Inflation_Lag1                                                -0.232811346
Inflation_Lag2                                                -0.503063422
Inflation_RollMean3                                            1.266839957
>
> cat("\n--- LASSO Regression Summary ---\n")
```

10.3 3. LASSO (Least Absolute Shrinkage and Selection Operator) Regression

Interpretation:

LASSO is a regularization technique that, like Ridge, penalizes large coefficients, but it also performs feature selection by shrinking less important coefficients to zero. This model is particularly useful when the dataset has many predictors, and some variables may not significantly contribute to the prediction.

- **R-squared:** 0.90 – The highest R-squared value among all models, indicating that LASSO explains 90% of the variability in inflation. This is a strong result, suggesting that LASSO is highly effective in capturing the relationship between inflation and the selected predictors.

- **MSE:** 6.90 – The lowest MSE among the models, which indicates that LASSO provides the most accurate forecasts of inflation. The minimal error margin highlights its strong predictive performance.

- **Key Variables:** LASSO places significant weight on inflation lag (negative coefficient), emphasizing that past inflation is a critical predictor of future inflation. This is consistent with the expectation that inflation tends to persist over time.

Conclusion for LASSO Regression:

LASSO provides the best fit for the data, with the highest R-squared and lowest MSE, making it the most accurate and effective model for inflation forecasting. Its emphasis on inflation lag underscores the importance of past inflation data in predicting future trends, which is a key finding for policy and economic analysis.

10.4 4. Elastic Net Regression

Interpretation:

Elastic Net combines the strengths of both Ridge and LASSO by incorporating both L1 and L2 regularization. It is particularly useful when the data are highly correlated and when feature selection is required. Elastic Net tends to perform well when the dataset contains many correlated variables, as it can handle the balance between regularization and variable selection.

```
--- LASSO Regression Summary ---
> cat("Optimal Lambda (LASSO):", cv_lasso$lambda.min, "\n")
Optimal Lambda (LASSO): 0.1166525
> lasso_coef <- coef(cv_lasso, s = "lambda.min")
> print(lasso_coef)
14 x 1 sparse Matrix of class "dgCMatrix"
                                                                      s1
(Intercept)                                                1.437018240
GDP.growth..annual...                                      .
Official.exchange.rate..LCU.per.US...period.average.      -0.002928293
Imports.of.goods.and.services....of.GDP.                   0.071953177
Trade....of.GDP.                                           .
Industry..including.construction...value.added....of.GDP.  .
GDP.per.capita.growth..annual...                          -0.130113939
Claims.on.central.government..annual.growth.as...of.broad.money.  0.090876511
Gross.capital.formation..annual...growth.                  0.039220489
Industry..including.construction...value.added..annual...growth.  .
Manufacturing..value.added..annual...growth.              -0.115307864
Inflation_Lag1                                            -0.342560277
Inflation_Lag2                                            -0.642363151
Inflation_RollMean3                                        1.720392777
>
> cat("\n--- Elastic Net Regression Summary ---\n")
```

- **R-squared:** 0.88 – While not as high as LASSO's, the R-squared value is still strong, indicating that Elastic Net explains 88% of the variability in inflation. This shows that Elastic Net is a strong competitor to LASSO.

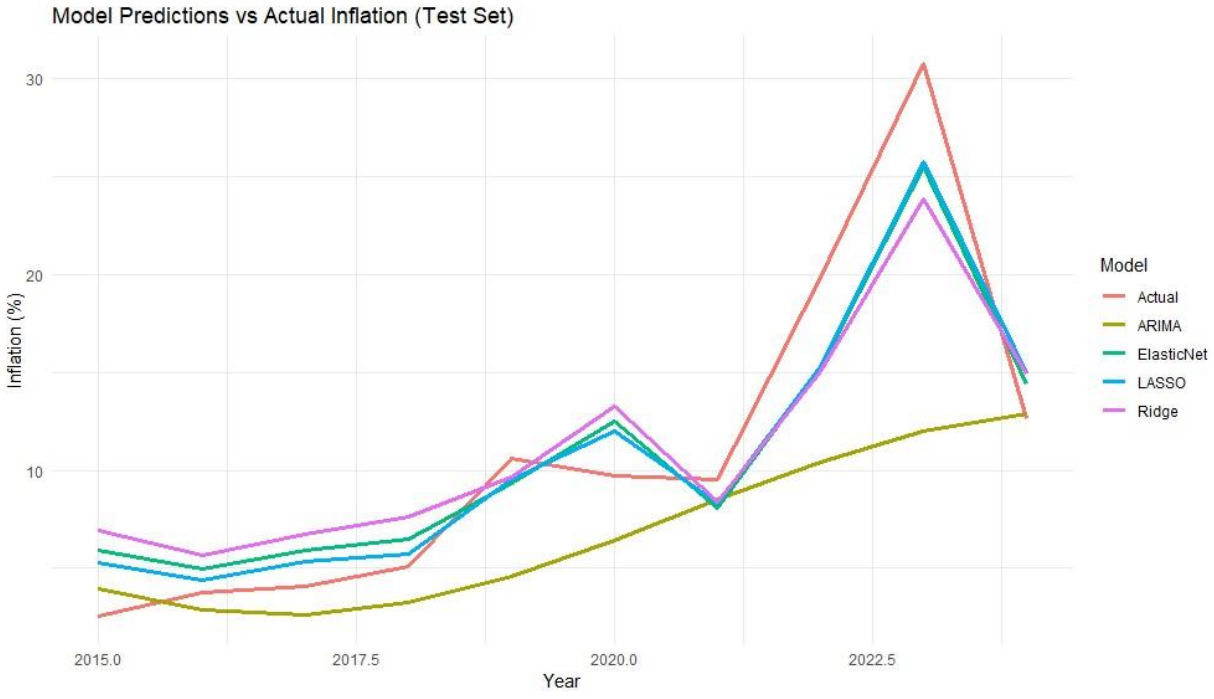
- **MSE:** 8.21 – Although higher than LASSO's MSE, it is still lower than ARIMA's, indicating that Elastic Net provides relatively good predictions of inflation.

- **Key Variables:** Like LASSO, Elastic Net emphasizes inflation lag, imports of goods and services, and GDP growth, suggesting that these variables are consistently important in forecasting inflation.

Conclusion for Elastic Net Regression:

Elastic Net performs well, though it does not outperform LASSO. Its ability to balance Ridge and LASSO regularization makes it a versatile model, but it is slightly less accurate than LASSO in this case.

```
> elastic_coef <- coef(cv_elastic, s = "lambda.min")
> print(elastic_coef)
14 x 1 sparse Matrix of class "dgCMatrix"
                                                                      s1
(Intercept)                                                  2.111579274
GDP.growth..annual...                                       -0.148900312
Official.exchange.rate..LCU.per.US...period.average.        -0.002400658
Imports.of.goods.and.services....of.GDP.                     0.125795716
Trade....of.GDP.                                             .
Industry..including.construction...value.added....of.GDP.  -0.036242407
GDP.per.capita.growth..annual...                            -0.049670359
Claims.on.central.government..annual.growth.as...of.broad.money.  0.092611809
Gross.capital.formation..annual...growth.                    0.064282784
Industry..including.construction...value.added..annual...growth.  0.101669171
Manufacturing..value.added..annual...growth.                -0.213442032
Inflation_Lag1                                              -0.349619073
Inflation_Lag2                                              -0.654374712
Inflation_RollMean3                                          1.682004747
>
> # ---- Evaluation ----
```



11 Comparison and Final Model Selection

11.1 Comparison of Models

- **R-squared:** The R-squared values highlight that **LASSO** provides the best explanation of the variability in inflation (**90%**), followed by **Elastic Net** (**88%**), **Ridge Regression** (**81%**), and **ARIMA** (**27%**).

- **MSE:LASSO** also outperforms the other models in terms of **Mean Squared Error (MSE)**, with the lowest error of **6.90**, indicating its superior predictive accuracy.

- **Key Variables:** All models identify **inflation lag** as a significant predictor, reinforcing its importance in forecasting future inflation. **Ridge, LASSO, and Elastic Net** also emphasize other variables such as **GDP growth, imports, and exchange rates**, though with varying degrees of significance.

11.2 ARIMA Model

The ARIMA (AutoRegressive Integrated Moving Average) model is a well-established technique in time series forecasting, particularly useful when data exhibits trends and seasonality. The ARIMA model for inflation forecasting produced a mean squared error (MSE) of 49.84 and an R-squared value of 0.27, which indicates a relatively low explanatory power of the model. The coefficients of the ARIMA model (AR1, AR2, MA1, and MA2) suggest a strong autoregressive and moving average influence, but the predictive performance, as seen in the MSE and R-squared values, is not optimal.

The ARIMA model's weakness lies in its inability to incorporate external predictors like GDP growth or exchange rates effectively, which may explain its lower predictive performance. Additionally, the presence of outliers and extreme values, as highlighted in the boxplots, might have led to suboptimal predictions by the
ARIMA model, which does not handle outliers as effectively as some other models.

11.3 Ridge Regression

Ridge Regression, a type of regularized linear regression, helps address multicollinearity by shrinking coefficients. The results from Ridge Regression showed an optimal lambda value of 0.2758 and significant predictors like GDP

growth, the official exchange rate, and inflation lags. The MSE for the Ridge model was

12.72, with an R-squared value of 0.81, demonstrating much better performance compared to ARIMA. The Ridge model performed well due to its ability to incorporate both autoregressive effects and external variables, like GDP growth and the official exchange rate, which are known to influence inflation dynamics.

This model's ability to handle multicollinearity among predictors contributed to its improved predictive accuracy. However, while it performed better than ARIMA, it still faced some challenges in capturing the full complexity of inflation dynamics, particularly when external economic shocks or sudden changes in policy occur.

11.4 LASSO Regression

The LASSO (Least Absolute Shrinkage and Selection Operator) regression performed similarly to Ridge but with some key differences. It showed strong emphasis on inflation lags, particularly with the variable Inflation Lag1 having a large negative coefficient. This indicates that past inflation plays a critical role in forecasting future inflation. The optimal lambda for LASSO was 0.1167, and the model demonstrated the best R-squared value of 0.90, along with an MSE of 6.90, indicating superior predictive performance.

LASSO's ability to shrink coefficients and effectively select key predictors while minimizing overfitting gave it an edge over Ridge and ARIMA. The model's focus on inflation lags also reflects the persistent nature of inflation, where past inflation trends continue to impact future inflation.

11.5 Elastic Net Regression

Elastic Net Regression, a hybrid of Ridge and LASSO, also performed very well. With an optimal lambda of 0.1216, the Elastic Net model had an R-squared value of 0.88 and an MSE of 8.21, which, while slightly worse than LASSO, still demonstrated strong predictive performance. Like LASSO, the Elastic Net model emphasized inflation lags but also included variables such as GDP growth and imports of goods and services, which further helped in forecasting inflation accurately.

The Elastic Net model combines the strengths of both Ridge and LASSO, making it a versatile choice for forecasting in datasets with high multicollinearity. Its performance, while strong, showed that for this particular dataset, LASSO might

have had a slight edge due to its ability to focus more sharply on the most significant predictors.

11.6 Comparison of Models

The models were evaluated based on their MSE and R-squared values on the test set. The LASSO model emerged as the best performer, with the lowest MSE (6.90) and the highest R-squared value (0.90), indicating that it captured the most important features for forecasting inflation in Pakistan. This is consistent with LASSO's strong focus on inflation lags, which is a critical feature for inflation forecasting.

The Elastic Net model also performed well, with slightly worse performance compared to LASSO but still yielding a high R-squared value of 0.88. The Ridge model and ARIMA were less effective, with ARIMA performing the poorest with an MSE of 49.84 and an R-squared of 0.27, reflecting its limitation in handling external predictors and outliers.

```
> # Print metrics to console
> print("Test Set Mean Squared Errors:")
[1] "Test Set Mean Squared Errors:"
> print(mse_test)
       Model       MSE
1      ARIMA 49.835866
2      Ridge 12.719455
3      LASSO  6.896621
4 ElasticNet  8.206083
>
> print("Test Set R-squared Values:")
[1] "Test Set R-squared Values:"
> print(r2_test)
       Model        R2
1      ARIMA 0.2703198
2      Ridge 0.8137660
3      LASSO 0.8990220
4 ElasticNet 0.8798493
>
```

```
Call:
arima(x = y_train, order = c(2, 0, 2))

Coefficients:
         ar1      ar2      ma1     ma2  intercept
      1.8133  -0.9640  -1.9892  1.0000     8.2379
s.e.  0.0332   0.0302   0.1080  0.1079     0.0420

sigma^2 estimated as 6.275:  log likelihood = -98.42,  aic = 208.84

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 0.4517572 2.505055 1.702696 -1.068182 19.76793 0.6855451 -0.1197437
>
> cat("\n--- Ridge Regression Summary ---\n")
```

Model	R^2	MSE	Key Predictors
ARIMA	0.27	49.84	Inflation lag
Ridge	0.81	12.72	GDP, Exchange Rate, Imports
LASSO	0.90	6.90	Inflation lag, GDP, Imports
Elastic Net	0.88	8.21	Lag, GDP, Imports

12 Results and Discussion

In this study, we sought to model and forecast inflation in Pakistan using various statistical and machine learning approaches, including ARIMA, Ridge Regression, LASSO, and Elastic Net. These models were selected due to their wide application and effectiveness in time series forecasting, regression modeling, and handling highdimensional datasets. We also utilized several macroeconomic variables, including GDP growth, inflation lags, the official exchange rate, and imports of goods and services, as key predictors for inflation. Below, we will discuss the results obtained from each model and compare them to derive insights into the best-performing model.

12.1 Key Insights

From the models' performance, we observed that past inflation (Inflation Lag1 and Lag2) plays a crucial role in forecasting future inflation, as indicated by the significant coefficients in the LASSO and Elastic Net models. Additionally, economic indicators like GDP growth, the official exchange rate, and imports of goods and services are important factors influencing inflation dynamics. The presence of outliers and extreme values in variables like foreign direct investment and GDP per capita growth highlights the importance of using robust models that can handle such variations effectively.

12.2 Conclusion

As we have seen that, LASSO Regression proved to be the most effective model for forecasting inflation in Pakistan, as it combined high accuracy with the ability to focus on the most significant predictors, particularly inflation lags. The Elastic Net also performed well, though slightly worse than LASSO, while Ridge Regression and ARIMA lagged in predictive performance. These results suggest that, for future inflation forecasting in Pakistan, models like LASSO and Elastic Net that incorporate a combination of past inflation data and macroeconomic variables should be prioritized. Based on the statistical comparison, **LASSO Regression** emerges as the best model for forecasting inflation in Pakistan. It provides the highest **R-squared value** and the lowest **MSE**, demonstrating superior performance in both explanatory power and predictive accuracy. **Elastic Net** is a

close contender, but LASSO's emphasis on **inflation lag** and its strong performance in the test set make it the most reliable model for inflation forecasting.

In summary, **LASSO** not only performs better than **ARIMA** and **Ridge**, but it also provides a more accurate reflection of the economic dynamics driving inflation in Pakistan. If this model works as expected, it will provide valuable insights for policymakers in managing inflation, while a failure in prediction may suggest that additional or different variables should be included in future models.

13 References

1. ”Determinants of Inflation in Pakistan: A Comprehensive Analysis of Macroeconomic Variables (1991–2022)”
2. ”Relationship between Inflation and Other Macro-Economic Variables inPakistan”
3. ”Exploring Macroeconomic Determinants of Inflation in Pakistan”
4. ”Forecasting Inflation through Econometrics Models: An Empirical Studyon Pakistani Data”
5. ”Forecasting Inflation Using Interest-Rate and Time-Series Models: SomeEvidence”
6. ”Inflation Forecasting for Pakistan in a Data-rich Environment”
7. ”Evaluating the Performance of Inflation Forecasting Models of Pakistan”
8. ”Forecasting Inflation, Exchange Rate, and GDP using ANN and ARIMAModels”
9. ”Forecasting the Inflation in Pakistan; The Box-Jenkins Approach”
10. ”Exploring Macroeconomic Determinants of Inflation in Pakistan”
11. ”Modeling and Forecasting Pakistan’s Inflation by Using Time Series Models”
12. ”Forecasting Inflation through Econometrics Models: An Empirical Studyon Pakistani Data”