INFLATION FORECASTING FOR PAKISTAN USING MACHINE LEARNING METHODS

Author

MUHAMMAD FAWAD INAM

Regn Number

00000204844

Supervisor

Dr. Hasan Sajid

ROBOTICS AND INTELLIGENT MACHINE ENGINEERING

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

AUGUST 2021

INFLATION FORECASTING FOR PAKISTAN USING MACHINE LEARNING METHODS

Author

MUHAMMAD FAWAD INAM

Regn Number

00000204844

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Robotics & Intelligent Machine Engineering

Thesis Supervisor:

Dr. Hasan Sajid

Thesis Supervisor's Signature: _____

ROBOTICS AND INTELLIGENT MACHINE ENGINEERING

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

AUGUST 2021

Thesis Acceptance Certificate

It is certified that the final copy of MS Thesis written by Muhammad Fawad Inam (Registration No. 00000204844), of Department of Robotics and Intelligent Machine Engineering (SMME) has been vetted by undersigned, found complete in all respects as per NUST statutes / regulations, is free from plagiarism, errors and mistakes and is accepted as a partial fulfilment for award of MS Degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in this dissertation.

Signature:_____

Date:_____

Dr. Hasan Sajid (Supervisor)

Signature HOD:_____

Date:_____

Signature Principal:_____

Date:_____

FORM TH-4

National University of Sciences & Technology

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: **Muhamad Fawad Inam, NUST Regn. no. 00000204844**. Titled: "Inflation prediction of Pakistan using machine learning methods" be accepted in partial fulfillment of the requirements for the award of **MS Robotics and Intelligent Machine Engineering** degree.

Examination Committee Members

1. Name: Dr. Hassan Sajid Signature:_____

2. Name: Dr. M. Jawad Khan Signature:_____

3. Name: Dr. Karamdad Kallu Signature:_____

Supervisor's name: Dr. Hasan Sajid Signature:_____
Date:_____

_____ _____
Head of Department Date

COUNTERSIGNED

Date: _____ Dean:_____

Declaration

I certify that this research work titled "*Inflation forecasting for Pakistan Using Machine Learning Methods"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

MUHAMMAD FAWAD INAM

00000204844

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Student

MUHAMMAD FAWAD INAM

00000204844

Signature of Supervisor

Copyright Statement

Acknowledgements

I am grateful to Allah (S.W.T) for granting me the patience and persistence to achieve this milestone in my academic career.

I would also like to thank my supervisor for their guidance during this exciting journey.

Lastly, I would like to thank my parents who have stood by me through thick and thin and supported me in this all journeys of life.

Abstract

Economic indicators are vital for any country to make better financial and policy making decisions. Well informed and timely decision are important keeping in view the pace of production, accessibility and supply chains. Economic indicators are dynamic and does not solely rely on financials of the market but also on supply chain, natural calamities and short- and long-term policies by the institutions. We are using statistical models and machine learning models to forecast inflation of Pakistan. A univariate approach is good for getting good forecast, but this is not suitable to capture turning points and identify the causation of inflation. This study involves multivariate approach to forecast consumer price index and to understand the relationship between CPI and other macroeconomic indicators.

Contents

Thesis Acceptance Certificate i

Declaration iii

Plagiarism Certificate (Turnitin Report) iv

Copyright Statement v

Acknowledgements 6

Abstract 7

1. Introduction 9

1.1. Overview of CPI 10

1.2. Problem Statement: 12

1.3. Objective: 12

1.4. Thesis Overview: 12

2. Literature Review: 12

3. Methodology: 15

3.1. Data: 15

3.1.1. Data Pre-Processing: 15

3.1.2. Feature Selection: 17

3.2. Proposed Method: 20

3.2.1. Sarimax: 21

3.2.2. Prophet model: 22

3.2.3. ANN without time series: 23

4. Results: 24

4.1.1. SARIMAX MODEL: 24

4.1.2. Prophet: 25

4.1.3. ANN: 26

5. Conclusion and Future Work: 26

References 27

1. Introduction

Inflation forecast has always been a center of attention for the governments and financial institutions as a guide in the formulation of monetary policy by the financial institutions around the world. Monetary policy decisions are based on the forecasted inflation of various models and other information provided by relevant economic indicators. Precise inflation forecasts are necessary for the central institutions to attain their targets and it also helps companies in decision making about the price and wage contracts. Current inflation levels which are the results of previous policies may give insufficient information. Forecasts that connect current developments and policies to future inflation can bridge this gap. It assumes that inflations forecasts are reliable. However, situations where quality of data is evolving, and structure is less stable, quantitative inflation forecast can provide insights on further development that needs to be matched with additional analysis beyond econometric relationships. If unexpected high inflation occurs, families that depends on bond income or pensions will be particularly affected with the costs over long duration. An unforeseen increase in inflation has the tendency to lower the labor wage and affects their buying power. Companies and families have to devote their time and efforts to decrease the currency holding and businesses to repeated adjustments in prices. Moreover, the cost of capital will increase by high inflation after tax payment. Therefore, such outcome is result of capital depreciation. [1]

Modeling inflation involves the selection of an inflation measure. In Pakistan, we have different measures of general trend in prices in the country. National statistics institute, Pakistan Bureau of Statistics is responsible for accumulating, compilation and broadcasting of the prices related data and indices. Such indices include GDP deflator, Consumer Price Index (CPI), Wholesale Price Index (WPI) and Sensitive Price Index (SPI). Within the basket of CPI. We also have and exclusion-based measures of core prices index and that is for Non-Food Non-Energy (NFNE) group. Another measure of core inflation for which PBS started publishing data is ’20 percent

trimmed core inflation. In calculating 20 percent trimmed core inflation, 10 percent of items showing extreme price changes each from top and bottom are excluded from the CPI basket.

SPI is the most frequently (weekly) available price index, but it covers only necessities and just 17 cities. GDP deflator is the most comprehensive one but is available less frequently (yearly). WPI does not cover the services. Core inflation is the one measure which SBP considers important in discussion in its flagship publications; but it is not the target inflation variable. So, we are left with CPI. Government of Pakistan announces annual inflation target which is for 12 months average of year-on-year change in CPI. In this study we will be using as our target variable.

Among all the price indices, SPI is the most frequent, as it is measured weekly, but it accumulates only necessities and seventeen cities. One which gives a broader understanding of economic dynamics is GDP deflator but only available annually. WPI only covers the prices of goods and not the services. Core inflation is the one measure which SBP considers important in discussion in its flagship publications; but it is not the target inflation variable [2].

1.1. Overview of CPI

Consumer Price Index (CPI) is used to recognize variations in the cost of purchasing a bucket of goods and services by a consumer. To compile the CPI, a price sample representative of the full set of goods and services available to consumers that takes account of the relative weight of each item in overall household expenditure is needed.

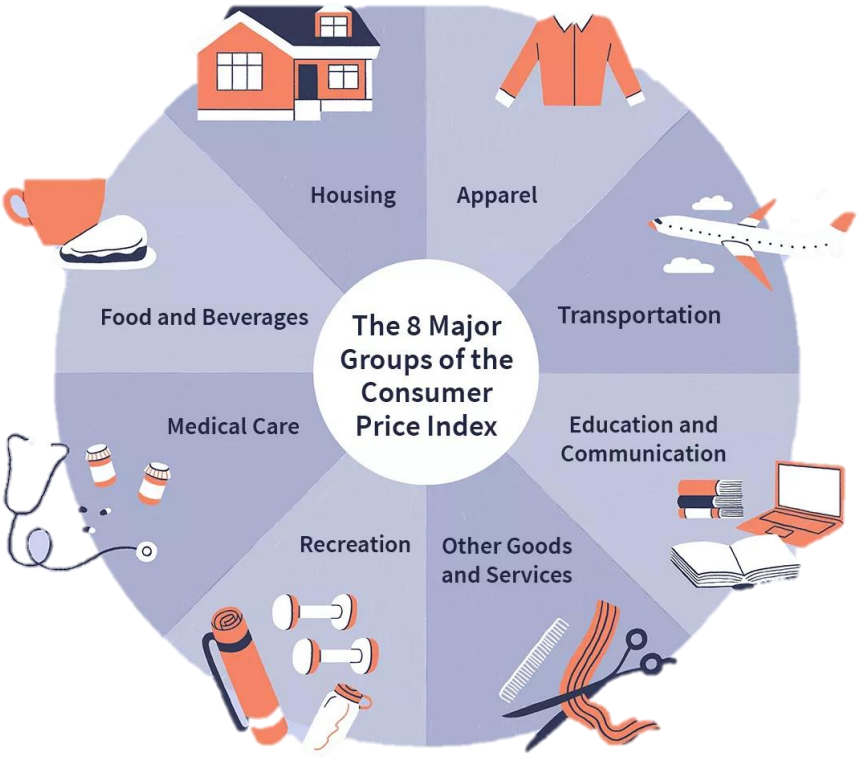


Figure 1: Goods & Services for CPI basket

This sample is defined, and weights are identified with respect to the information extracted by the financial or statistical institutions, as well as additional sources. The current CPI, which has its base year as 2007-2008, is calculated using 148048 price quotes every month corresponding to 487 products [3]. Following is how the CPI is calculated given the weights of items are defined for the basket.

$$CPI_t = \frac{C_t}{C_0} * 100$$

CPI_t = consumer price index in current period
C_t = cost of market basket in current period
C_0 = cost of market basket in base period

Gives the consumer price index with respect to the base period which is 2008 in case of Pakistan.

1.2. Problem Statement:

Inflation is one of the most looked after indicator in any economy as it defines the prices of goods and services available in the region. CPI is a metric used for inflation which is calculated using the price basket of goods and services. Reliable forecasting of inflation metric can help make better decisions and policies.

1.3. Objective:

The main objective of this study is to provide methods and to forecast the metric used for inflation i.e., CPI using some high-level indicators instead of using the price basket method. Doing so to provide tool for better economic forecasting in scarce data environment.

1.4. Thesis Overview:

In this work, section 2 explains the work done by previous researchers and methods they have used with respect to forecasting inflation. Section 3 contains the methodology and implementation including the dataset, data pre-processing and complete workflow. Section 4 includes the results acquired after implementing the methods to forecast. Section 5 consists of discussion of complete work and section 6 describes all the possible future work which can be held in this domain.

2. Literature Review:

There is substantial and increasing literature on inflation forecasting in emerging economies. Analogous to the insights for advanced economies, changes in price of imports, money growth, inflation expectations, nominal exchange rates, and exogenous supply shocks, especially to food and oil prices, are identified as main determinants of inflation in emerging economies [4].

Many empirical studies find standard economic relationship to hold. Estimates of money demand functions mostly find money demand to be determined by measures of opportunity

costs and activity [5]. Similarly, inflation is influenced by changes in money supply, measures of aggregate demand or output, interest rate and import prices [6]. However, many studies find such relationships to hold in a cointegration framework, some failed to identify cointegration which could imply to the structural breaks in particular samples [7]. There seem to be no or only little exchange rate between local prices [8].

Following table gives a detail overview of the previous work being carried out regarding inflation in Pakistan.

Table 1: Past Work

Ref	Approach	Dependent Variable	Regressors	Sample Period	Findings
[9]	VAR		CPI inflation, WPI inflation, M2, oil prices	1988:1 to 2003:9	Little exchange rate pass through to domestic inflation
[8]	VAR	CPI, WPI	USD exchange rate, foreign price index	1982-2003	There is no exchange rate pass-through to domestic prices
[10]	OLS, cointegration analysis	M1, M2	Industrial Production, Interbank call money rate, CPI inflation	1972: I to 1996: I	Inflation is a better measure of opportunity cost than interest rate.

[6]	Single equation, cointegration test, 2-equation model with 2SLS	CPI and Exchange Rate	Exchange rate, import prices, world prices, money supply, GDP, forex reserves	1982: II to 1996: IV	CPI reacts to changes in import prices and money supply. Exchange rate responds to domestic and world prices.
[11]	OLS	CPI	Bond Money, GDP growth, share of service sector, public debt, import prices	1972-1992	Inflation results from money growth and structural factors.
[12]	OLS of distributed lag specification	CPI	M1, industrial production, interest rate, foreign interest rate, import prices	1970: I to 1987: IV	All variables explain inflation except M1
[13]	OLS	WPI,CPI,GNP deflator, and absorption deflator	Real GNP growth, lagged inflation, growth rate of unit value of imports, growth rate of M1/M2	1960 to 1988	Inflation is determined by real GNP growth, unit value of import growth, nominal money growth and lagged inflation.

3. Methodology:

3.1. Data:

While developing a forecasting model for Pakistan, there could arise few problems collecting and utilizing data. One of the major problem Pakistan does not used to have a data rich environment or data centric approach previously to identify the problems. So, with less data there were few factors to measure their relationship with inflation. Pakistan Bureau of Statistics (PBS) is the institute responsible for the collection and compilation of data. But PBS was not used as a single source of information since it mostly contains reports and statistics for each month or quarter or year separately. Data was accumulated using various sources which includes World Bank, State Bank of Pakistan, Pakistan Bureau of Statistics, IMF and few third-party data economic data portals like CIEC and Trading Economics. It was tried to collect as much data possible without compromising the quality of data. Therefore, data from 2001 to 2021 was acquired as these years were more data rich and give a more accurate data than previous years.

Table 2: Summary of Dataset

Total Features	26
Data Frequency	Monthly
Feature Type	Numerical
Total Records	241
Data characteristic	Time series

3.1.1. Data Pre-Processing:

Missing values are one of the major and common problems in the datasets. Solving missing values can vary from problem to problem. Missing values can be handled using replacing null

with the zero or with mean values. For time series problem it seems reasonable to use linear interpolation.

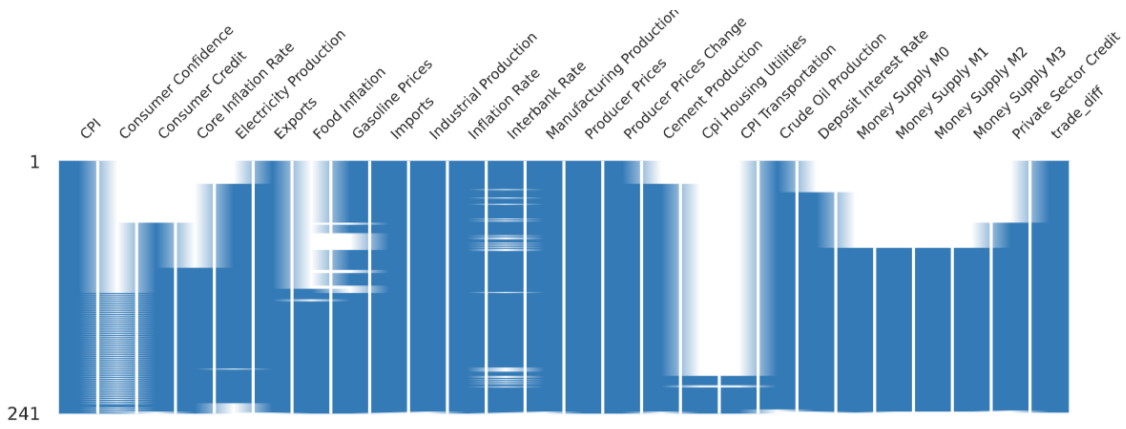


Figure 2: Missing Values

As it can be seen in Figure 2 that most of the predictors have missing values even after collecting the most recent data. Data was refined using three methods in order to maintain the data quality for the model.

1. Predictors with very high missing values were dropped as they will not be able to give a clear understanding of the results.
2. Linear interpolation was used for few features as it gives more sense when dealing with time series. Because these feature does show a gradual increasing or decreasing pattern which makes linear interpolation suitable. Example, Gasoline prices.
3. Some features with missing values do not show increasing or decreasing trend but their values were correlated with other available features. Feature with highest correlation to the column of missing values was selected, corresponding values were observed, and missing values were replaced by the average of the corresponding values.

3.1.2. Feature Selection:

In every multivariate timeseries dataset there are features that contribute to forecasting the target feature but there are also some features that contribute less or including them would make the model fitting more complex affecting the results. Initially target variable was plotted to see the trend.

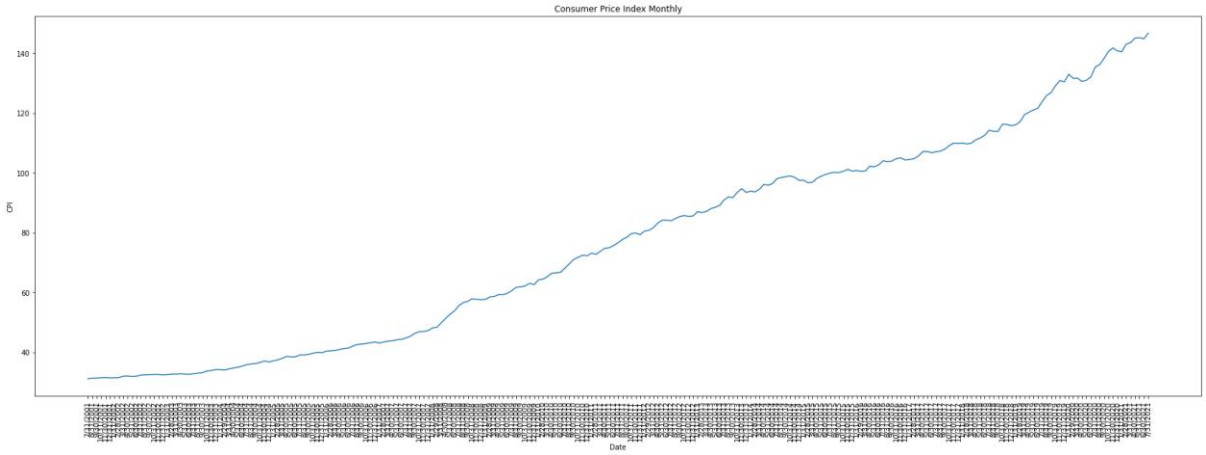
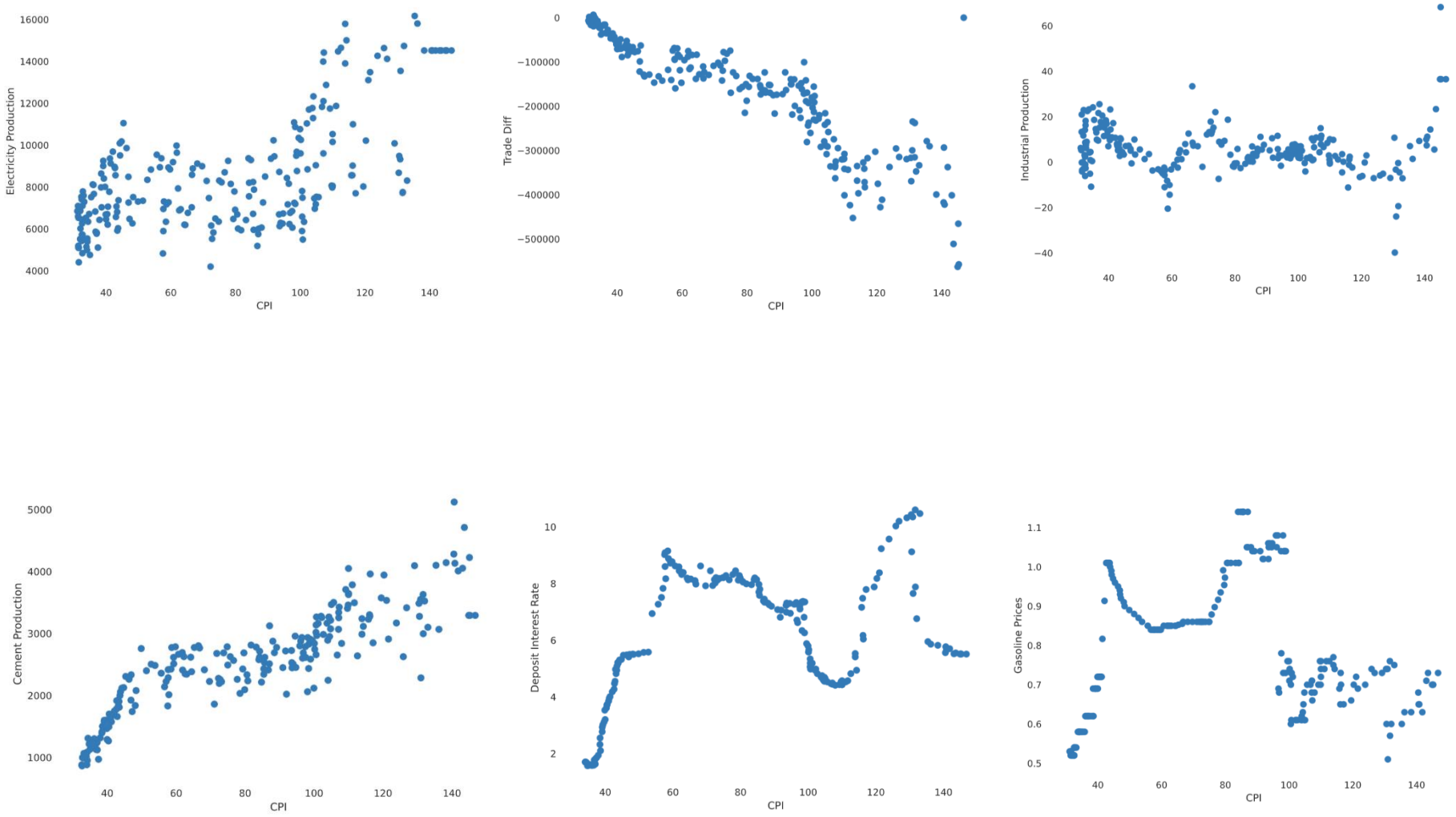


Figure 3: Monthly CPI 2001-2021

This provides an increasing trend over time. Further we used feature interaction plots (Figure 4) to identify the behavior of predictors with the target variable for feature selection. To be more concrete Pearson Correlation was calculated between all the features, plotted a heatmap (Figure 5) and removed the features which are highly correlated with each other.

Figure 4: Interaction Plots of CPI (target variable) with few predictors



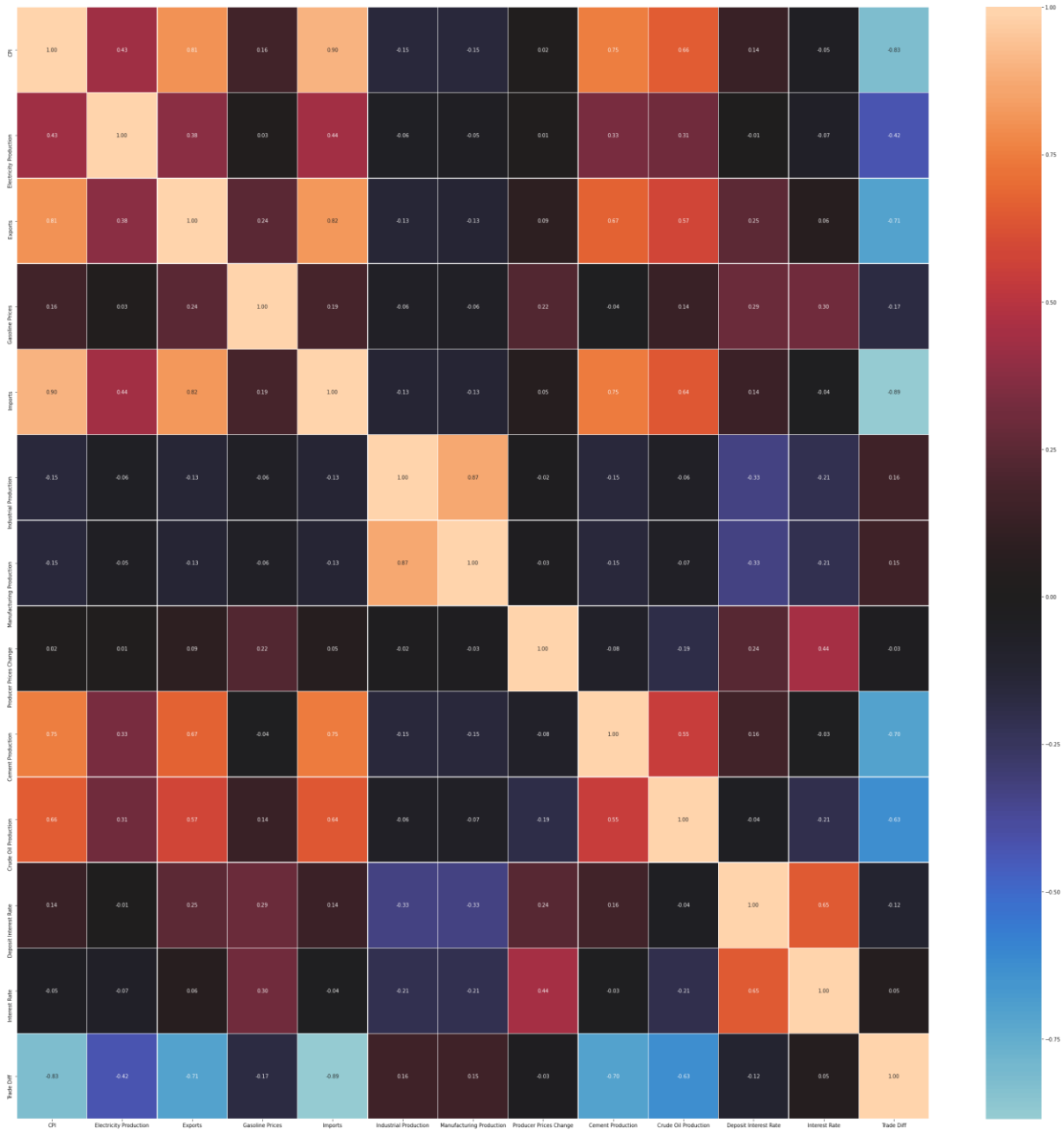


Figure 5: Correlation Plot

Features having absolute correlation values greater than 0.3 with the target variable, were selected to be used for modeling. Electricity production, gasoline prices, manufacturing production, deposit interest rate, crude oil production, trade difference seem to be of high importance while looking the correlation values.

3.2. Proposed Method:

With the preprocessing of all the data and feature selection comes the behavior of the target variable. In time series problems there could exist some seasonality in the data. Figure 6 gives a better understanding of trend and seasonality of the target variable.

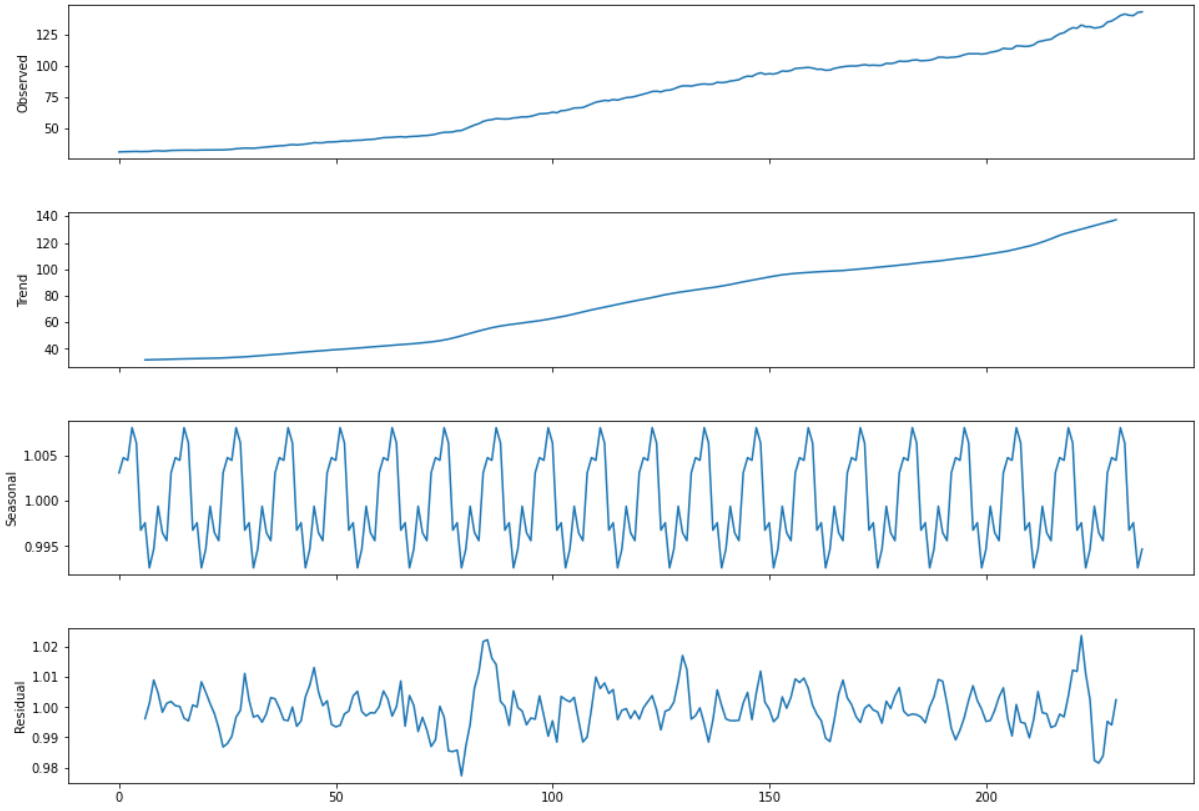


Figure 6: Trend of CPI

Further the data was investigated for the autocorrelation. If the current value depends on the previous value and how much it is likely to define the next value. Figure 7 describes the autocorrelation and partial correlation with lag of 12.

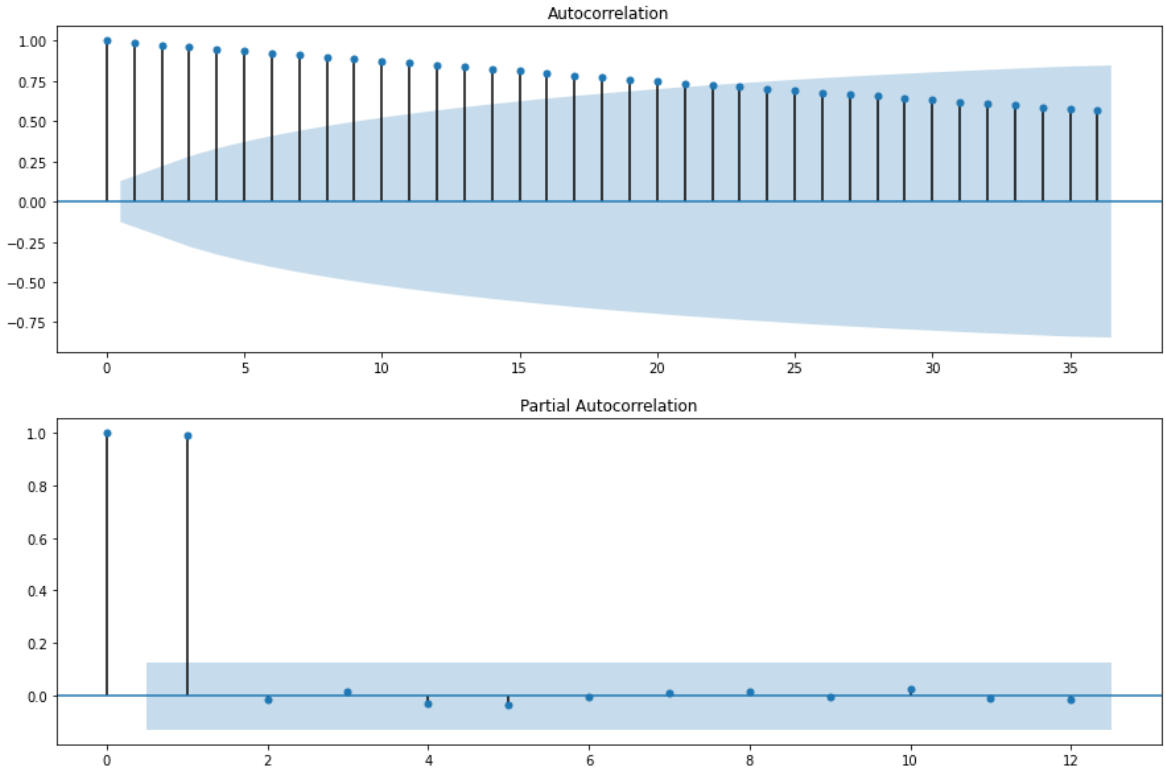


Figure 7: ACF & PACF Plots

3.2.1. Sarimax:

Auto-Regressive Integrated Moving Average (ARIMA) is frequently used for timeseries forecasting but the problem with ARIMA is that it's a univariate forecasting model which cannot incorporate multiple features and seasonality in the timeseries data. To tackle such problems several variations of ARIMA have been introduced.

Seasonal Auto Regressive Integrated Moving Average (SARIMA) is an extension of ARIMA to handle seasonality within the time series data, but even this had the shortcoming of being a univariate model. To solve multivariate forecasting problems that has seasonality effects as well as multivariate features SARIMAX was introduced. SARIMAX handles seasonality which ARIMA could not and accommodates exogenous variables. SARIMAX function can be defined by Eq. 1

$$\Theta(L)^p \theta(L^s)^p \Delta^d \Delta_s^D y_t = \phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \epsilon_t + \sum_{i=1}^{n} \beta_i x_t^i \qquad Eq.\ 1$$

Where p, P, q, Q, d and D are the non-seasonal and seasonal AR, MA and differencing orders, respectively, while L is lag operator and β is a constant.

The *statsmodel* python library function SARIMAX is used to fit and forecast our timeseries data with the configuration in Table 3

Table 3 SARIMAX Parameters

PARAMETERS	p	P	q	Q	d	D	s
VALUES	0	0	1	0	0	0	1

3.2.2. Prophet model:

The forecasting model used by Prophet, a python package developed by Facebook, is based on an additive model and uses decomposable timeseries forecasting model. This model was initially built for business forecasting purposes having strong seasonal effects but due to its ability to incorporate seasonality, trend, and multivariate features it can be customized for use in weather time series forecasting as well. The *Prophet* model is shown in Eq.2:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \qquad \textit{Eq.2}$$

In the above equation $g(t)$, $s(t)$ and $h(t)$ are the trend, seasonality, and holidays function respectively. The term ε_t represents characteristic changes at time-step 't', that are not accommodated by the model. Since, solar irradiance by any type of holidays, so, $h(t)$ function is not used.

The Prophet forecasting model uses Analyst-in-the-loop model shown in Figure 6, because often analysts have domain knowledge about the variable being forecasted but have little statistical knowledge and there are there are several places in the model specifications where analysts can apply their expertise and external knowledge without requiring deep understanding of the underlying statistics.

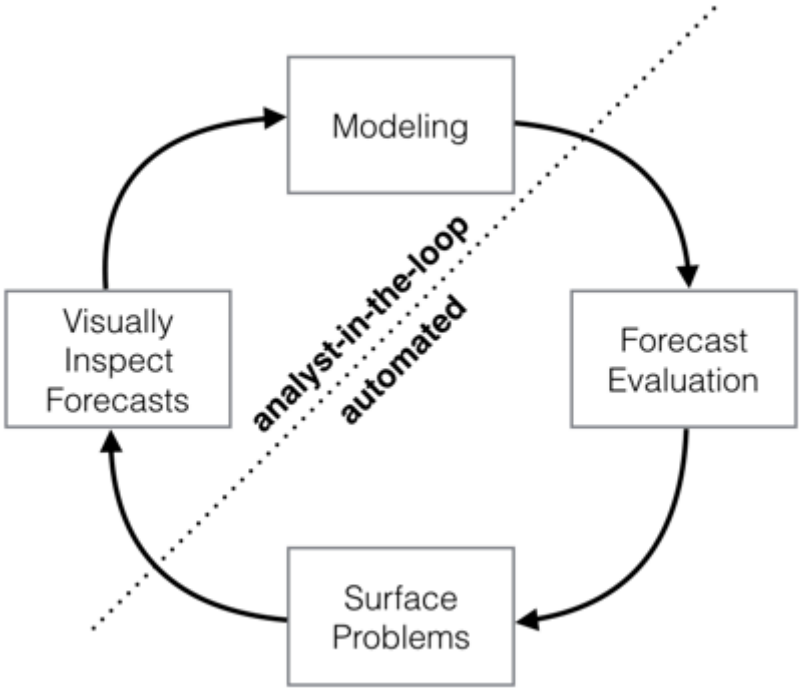


Figure 8: Analyst-in-the-loop model

3.2.3. ANN without time series:

An artificial neural network (ANN) is a system in Machine Learning built to mimic on the way the human brain analyzes and processes information, having self-learning capability to solve problems that would, otherwise, be impossible or difficult by human or statistical standards and depends on data experiences to learn and predict results as humans do.

ANNs consist of three main units namely Input Layer, Hidden Layers and Output Layer. The hidden layer is where the learning and processing happens. The hidden layer consists of neurons that are the base processing units of each layer. The hidden layers take data from the input layer and learns depending on the type of learning method chosen for each hidden layer and forwards the result to the output layer. A simple architecture of Artificial Neural Networks is given in Figure 7.

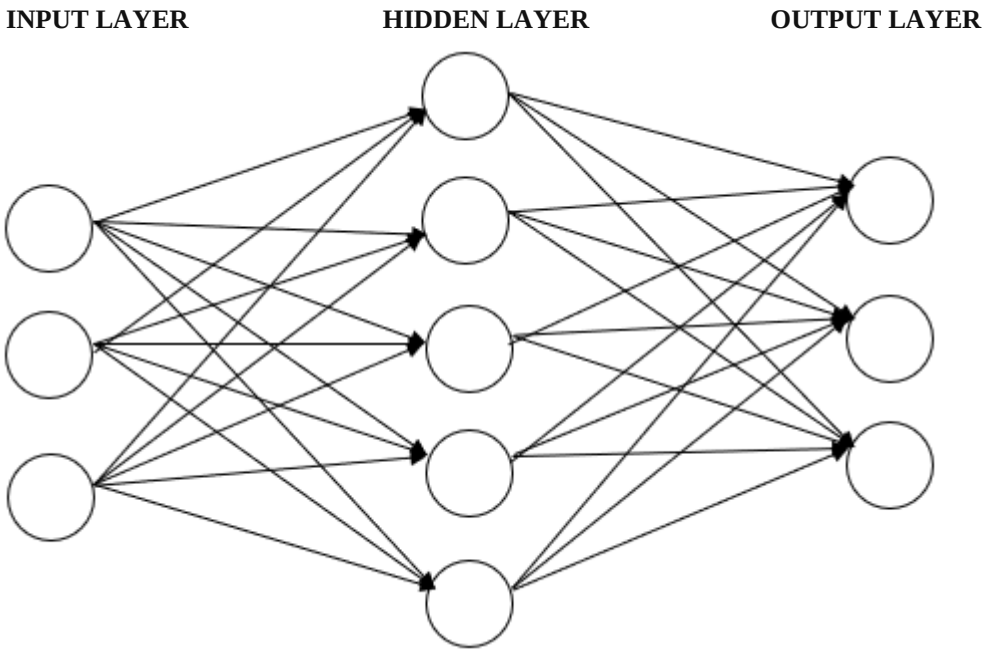


Figure 9: Artificial Neural Network

The ANN configuration used in this study to perform time series forecasting of CPI.

4. Results:

This section discusses the results generated by the different iterations of models.

4.1.1. SARIMAX MODEL:

Four different iterations were used for the Sarimax where different moving average and autoregressive parameters were used in order to see the impact of it.

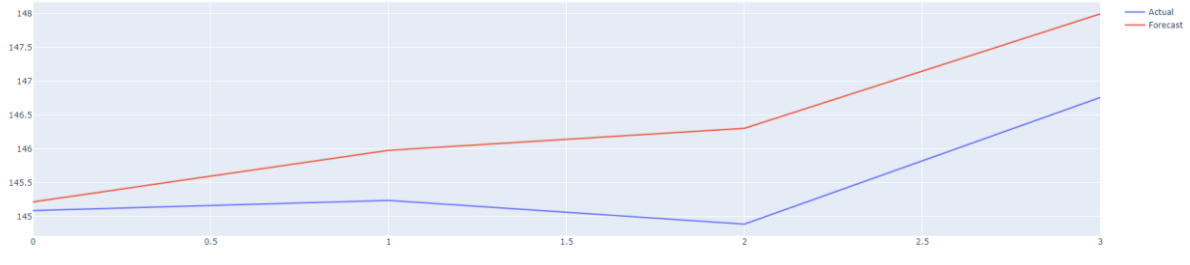


Figure 10: Actual and Forecast Results, MA = 4 , AR = 4

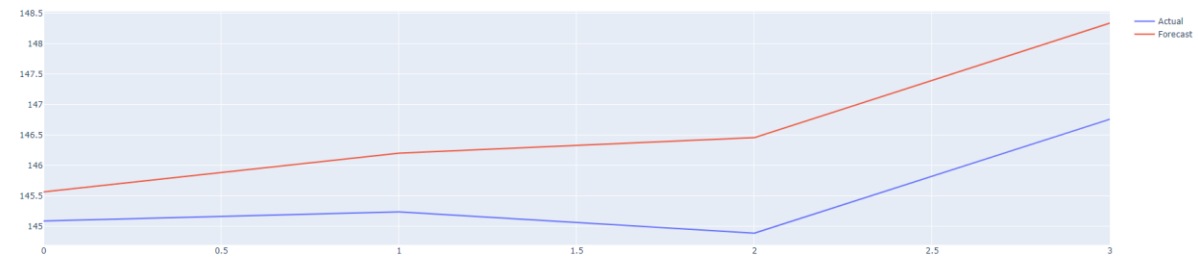


Figure 11: Actual and Forecast Results, MA = 8 , AR = 4

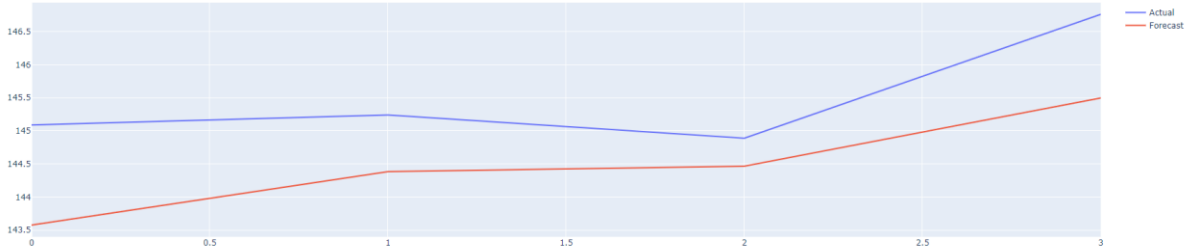


Figure 12: Actual and Forecast Results, MA = 2 , AR = 1

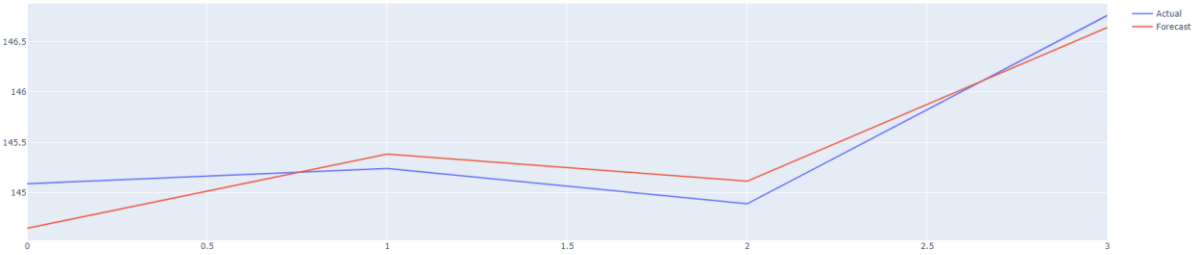


Figure 13: Actual and Forecast Results, MA = 1 , AR = 1

4.1.2. Prophet:

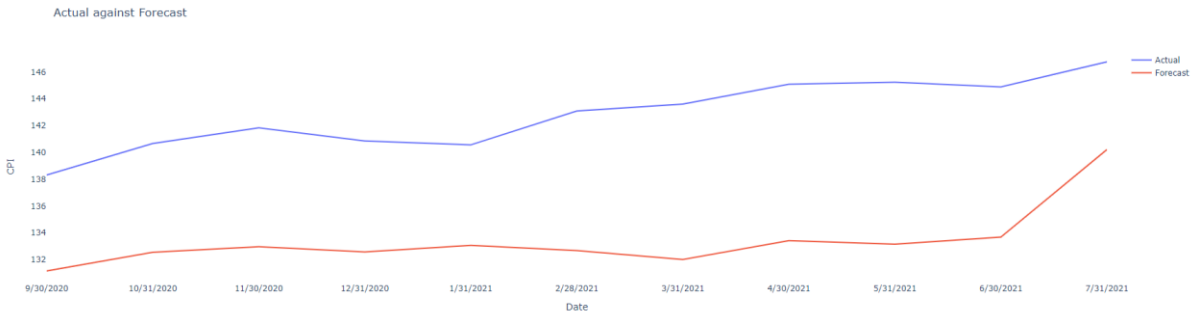


Figure 14: Actual and Forecast results, Prophet additive Model

4.1.3. ANN:

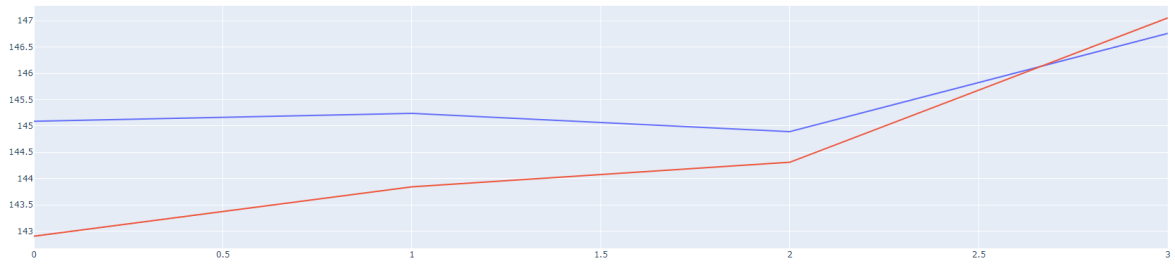


Figure 15: ANN results with 3 connected layers of 10 neurons

Table 4: Comparison

Model	R2	RMSE	MSE	MAE
SARIMAX	0.87	0.26	0.07	0.23
Prophet (Additive model)	-14.30	9.60	92.26	9.40
ANN	0.13	4.69	42.66	4.30

5. Conclusion and Future Work:

We were able to identify variables affecting the inflation Pakistan and created exogenous models. Statistical model Sarimax performed better than prophet and ANN. While ANN appear to perform good given the size of the data. Machine learning model will be able to perform better with increased number of records and better processed features.

More features can be identified to verify their behavior on inflation metric. LSTM, CNN or Deep NN can be used to identify rich patterns of variables with the target variable.

References

[1] C. J. L. Yellen, "Inflation dynamics and monetary policy [speech]," 24 September 2015. [Online]. Available: https://www.federalreserve.gov/newsevents/speech/yellen20150924a.htm.

[2] M. J. M. Muhammad Nadeem Hanif, "Evaluating the performance of inflation forecasting models of Pakistan," *SBP Research Bulletin,* vol. 11, no. 1, p. 43, 2015.

[3] P. B. o. Statistics, "Press Release on Consumer Price Index Inflation," Islamabad, 2019.

[4] M. Bokil and A. Schimmepfennig, "Three attempts at inflation forecasting in Pakistan," *The Pakistan development review,* vol. 45, no. 3, pp. 341-368, 2006.

[5] Tariq, S. Muhammad and K. Mathews, "The demand for simple sum and Divisia Money aggregates for Pakistan: A Cointegration approach," *The Pakistan Development Review,* vol. 36, no. 3, pp. 275-91, 1997.

[6] Ahmad, Eatzaz and S. A. Ali, "Relationship between exchange rate and inflation," *Pakistan Economic and social Review,* vol. 37, no. 2, pp. 139-54, 1999.

[7] Shamsuddin, A. F. M. and R. A. Holmes, "Cointegration test of the monetary theory of inflation and forecasting accuracy of the univariate and vector ARMA Models of Inflation," *Journal of Economic Studies,* vol. 24, no. 5, pp. 294-306, 1997.

[8] Choudhri, E. U. and M. S. Khan, "The Exchange rate and consumer prices in Pakistan: is Rupee devaluation inflationary," *The Pakistan development review,* vol. 41, no. 2, pp. 107-20, 2002.

[9] Hyder, Zulfiqar and S. Shah, "Exchange Rate Pass-through to domestic prices in Pakistan (working paper No.5)," State Bank of Pakistan, 2004.

[10] Ahmad, Eatzaz and M. Munir, "An analysis of Money and demand in Pakistan," *Pakistan economic and social review,* vol. 38, no. 1, pp. 47-67, 2000.

[11] Chaudhary, M. Aslam and N. Ahmad, "Sources and Impacts of inflation in Pakistan," *Pakistan Economic and Social Review,* vol. 34, no. 1, pp. 21-39, 1996.

[12] Dhakal, Dharmendra and M. Kandil, "The inflationary Experience of six developing countries in Asia: An investigation of underlying determinants," *Applied Economics,* vol. 25, pp. 413-25, 1993.

[13] Ahmad, Eatzaz and H. Ram, "Foreign price shocks and inflation in Pakistan: A monetarist approach," *Pakistan Economic and Social Review,* vol. 29, no. 1, pp. 1-20, 1991.