A Deep Learning Approach to Forecast the Canadian Consumer Price Index (CPI) Using Encoder-Decoder Attention Mechanisms with Teacher Forcing Techniques

A Major Research Paper

presented to the Toronto Metropolitan University

in partial fulfillment of the requirements for the degree of

Master of Science in Data Science and Analytics

Department of Engineering

MRP Supervisor:

Dr. Aliaa Alnaggar

Second Reader:

Dr. Ceni Babaoglu

Candidate:

Maame Cobbinah,

BSc Economics and Statistics, University of Victoria, 2017

2022/2023

Author's Declaration for Electronic Submission of a Major Research Project (MRP)

I hereby declare that I am the sole author of this Major Research Paper. This is a true copy of the MRP, including any required final revisions. I authorize Toronto Metropolitan University to lend this MRP to other institutions or individuals for the purpose of scholarly research. I further authorize Toronto Metropolitan University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my MRP may be made electronically available to the public.

Maame Cobbinah

Abstract

This paper proposes a novel deep learning approach that utilizes a RNN encoder-decoder Attention model with Teacher-Force technique to forecast the Canadian Consumer Price Index (CPI). Although this method is quite complex in its implementation, it performs well with macroeconomic data, such as the CPI, due to its ability to model long-term dependencies in inflationary variables and reduce forecasting errors. The experimental part of this paper focuses on evaluating the performance of the proposed method, along with traditional methods, supervised learning techniques, and other deep learning approaches. The results show that the RNN Encoder-Decoder Attention Model with Teacher Force Technique outperforms existing methodologies in the literature, generating lower errors and improving R-squared values across all CPI indicators.

Keywords: *Consumer Price Index, macroeconomic data, Forecasting, RNN Encoder-Decoder, Deep learning, Attention Models, Teacher Force Technique*

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

List of Tables

List of Figures

1 Introduction

The development of CPI (Consumer Price Index) forecasting in macroeconomics has become crucial due to its significance in informing monetary policy decisions and guiding economic planning (Nyoni, 2019) (1). Accurate CPI forecasting helps policymakers and economists anticipate inflationary pressures, understand the dynamics of price changes, and assess the overall health of an economy. By providing insights into inflation trends, CPI forecasting enables policymakers to take proactive measures, such as adjusting interest rates and implementing appropriate fiscal policies, to maintain price stability and support sustainable economic growth. Concurrently, the growth in deep learning techniques has improved ability to forecast into the future. The integration of encoder-decoder attention mechanisms in Recurrent Neural Networks (RNNs) methods have shown promise in improving the accuracy and effectiveness of CPI forecasting models. The aim of this research paper is to explore the efficacy of this deep learning approach in predicting the Canadian Consumer Price Index, contributing to the existing literature on forecasting methods and providing valuable insights for policymakers and economists.

1.1 Background

The CPI is an economic indicator that measures the average cost variation in a diverse range of goods and services purchased by households, businesses, and other entities (Statistics Canada, 2023) (2). It focuses specifically on items considered as consumer goods or services and is typically associated with retail prices. The CPI excludes expenditures, such as income taxes, charitable donations, pension contributions, and consumer savings and investments (Statistics Canada, 2023) (2).

On a monthly basis Statistics Canada gathers prices across a spectrum of good and services. The CPI is then determined by assigning weights to each item comparing the total expenditure on a specific item to the overall spending in Canada (Statistics Canada, 2023) (2). These weights play a crucial role in assessing the impact of price changes on the overall consumer budget.

The CPI is anchored to a time base, or base period, which represents the period when the index is assigned a value of 100. Currently, the CPI time base in Canada is set at 2002. When referencing

specific index levels, it is customary to include the base period to provide context. For example, the CPI All-items for Canada in January 2020 was reported as 136.8 (2002=100), indicating that consumer prices were 36.8 % higher in January 2020 compared to the base period of 2002. To ensure an accurate representation of short-term inflation trends, the CPI incorporates adjustments for anticipated seasonal price fluctuations when comparing month-to-month changes in the index.

Recently, the Bank of Canada has relied on the CPI as a primary tool for quantitative tightening. This has resulted in several interest rate hikes, affecting the fees associated with mortgages, loans, and other forms of borrowing for Canadians. The bank's goal is to ultimately restore price stability and achieve balance in its balance sheet.

The CPI is a good inflation indicator as it calculated based on the representative basket of goods. This basket is designed to reflect the consumption patterns of the average consumer, making it a representative sample of consumer spending. Secondly, the frequency of data collection and measurement for CPI helps to provide timely updates on inflation trends. This frequency allows policymakers and economists to monitor changes in consumer prices and make informed decisions accordingly.

Despite its usefulness, the CPI has some challenges in terms of measurement. For example, quality adjustments is very hard to measure. Although, the CPI aims to measure price changes while holding the quality of goods and services constant, the improvements in the quality of products over time, such as advancements in technology or changes in product features, are not always adequately accounted for in the index. This can result in an underestimation of the true cost of living. Secondly, it is hard to take geographical variation into account when looking at the aggregated index value. The CPI represents an average measure of inflation for the entire country. It may not capture regional variations in prices, where different areas experience different rates of inflation. For example, the cost of transporting food to Northern Canada turns to be higher and hence food prices are relatively higher in these areas compared to the rest of Canada. The aggregated gross value may hide these important details .

1.2 Research Question

The core of this MRP centers around the fundamental research question: Can the RNN encoder-decoder attention mechanisms effectively enhance the accuracy and effectiveness of CPI forecasting? The exploration of this question seeks to shed light on whether RNN encoder-decoder attention mechanisms stand as a promising avenue for improving the accuracy and reliability of CPI forecasts, ultimately contributing to the advancement of predictive modeling in economics.

1.3 Independent/Dependent Variables

The list of all the available fields in the dataset is given in Table 1.

2   Literature Review

The following section examines previous works in time series forecasting in particular CPI and other economic/financial forecasting tasks and compares the performance of different models and techniques in these tasks.

2.1   Traditional Forecasting Tasks

Literature on CPI forecasting has mainly focused on traditional statistical techniques and basic machine learning methods. Atkeson & Ohanian (2001) (3), for example, examined such traditional techniques evaluating the effectiveness of Phillips curve-based models, specifically the NAIRU models, in forecasting CPI inflation. Through simulated exercises, it concluded that these models, including NAIRU Phillips curves, are not reliable tools for accurately predicting inflation, and alternative approaches should be explored to improve forecasting.

Similarly, Dondong (2010) (4) presented a simple Autoregressive Integrated Moving Average model. Dongdong's findings suggest that the ARIMA (12,1,12) model is effective in predicting the monthly CPI of China. However, this model suffers from data stationarity issues. Some CPI data often exhibit non-stationary behavior, such as trends or seasonal patterns. Although ARIMA performs better than Phillips curves in many cases it does not account for additional data transformations or differencing and hence can generate inaccurate output.

2.2   Supervised Learning Forecasting Tasks

As time progressed, supervised learning techniques emerged as better alternatives for forecasting CPI and other macroeconomic indicators. Medeiros et al. (2021) (5) introduced the random forest and lasso regressor model as an alternative to forecast CPI. The results showed that the random forest regressor outperformed traditional techniques such as VAR and ARIMA across all evaluation metrics. Wang et. al (2012) (6) proposed the Support Vector regressor as a better alternative for forecasting the Consumer Price Index (CPI). Their study concluded that using support vector regression (SVR) for CPI prediction is effective, as the nonlinear mapping capability of SVR allows for a better representation of the complex relationship between financial indicators and future CPI.

However, the model did not incorporate time-varying weights, highlighting the need to explore alternative optimization methods for such cases.

2.3 Hybrid Models

Hybrid models have also become increasingly popular. Liu et al. (2023) (7) proposed a hybrid model for CPI forecasting that combines Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Empirical Mode Decomposition (EMD), Hierarchical Agglomerative (HAC), and the Independent Component Analysis (ICA). The model decomposes CPI data into adaptive modes using CEEMDAN and clusters them with HAC to reduce dimensions. ICA is then employed to extract hidden information, and NAR is used for forecasting each independent component. The hybrid model outperforms benchmark models such as the ARIMA and NAR models, showing lower MAE and RMSE values and higher R^2 scores.

2.4 Deep Learning Techniques

Recently, deep learning techniques have emerged as better alternatives for forecasting. This due to their ability to handle complexity in time series data, perform transfer learning, and scalability. Bhardwaj (2022) (8) proposed an Artificial Neural Network (ANN) method for forecasting GDP per capita, achieving an impressive R^2 score of 0.9980 and a mean absolute error of 1.18%. Nakamura (2005) (9) compared a Neural Network (NN) model to autoregression (AR) models for CPI inflation forecasting, with the NN model outperforming in short-term forecasts. However, for longer forecast horizons, the NN model's performance becomes similar to the AR model. Barkan et al. (2022) (10) utilized a hierarchical recurrent neural network (HRNN) model to improve the prediction of disaggregated CPI inflation components, outperforming baseline models with an RMSE of 0.78 and correlation coefficients of 0.24 and 0.29. Theoharidi et al. (2023) (11) proposed a hybrid deep learning model, VAE-ConvLSTM, for inflation forecasting. It combined Variational Autoencoders (VAEs) and Convolutional LSTM Networks (ConvLSTM) to capture nonlinear patterns in inflation time series data. The model outperformed benchmark models, demonstrating superior consistency.

2.5 Transformer and Attention Models

This MRP however focuses on attention models. The attention methodology was introduced by Vaswani et al (2017) (12) in their paper "Attention is All You Need" . The model showcases the effectiveness of an attention Transformer model in language translation tasks. It achieved state-of-the-art results with improved efficiency, modeling long-range dependencies without recurrent or convolutional operations.

Abbasimehr & Paki (2022) (13) builds on this concept by applying attention to several time series datasets. Their study applied a hybrid LSTM and multi-head attention model to 16 public datasets. The model utilized a single-step ahead forecasting strategy with a selected lag size and optimized hyperparameters.The hybrid LSTM and multi-head attention model outperformed benchmarks, ranking first among the methods used.

Similarly, Chen (2022) (14) incorporated attention mechanisms based on time and sequence characteristics to capture important influencing factors for financial markets. Their proposed Deep Neural Network model (MALSTM) outperformed classical time series models (ARIMA) and LSTM models, improving MAE from 8.96% to 4.26%. Li et al. (2018) (15) also introduced an attention-based multi-input LSTM model for stock price prediction. Their study focused on China's CSI-300 index and incorporated a dual-stage attention mechanism using a mainstream factor and three auxiliary factors. The MI-LSTM model achieved a significant 9.96% improvement in mean square error compared to the LSTM model, demonstrating its effectiveness in stock price prediction. Finally, Du et al. (2020) (16) proposed an effective approach for forecasting multi-step time series values using an attention-based encoder-decoder framework. The framework incorporates a bidirectional Long Short-Term Memory (Bi-LSTM) as the encoder, an LSTM as the decoder, and a temporal attention context layer. The authors evaluated their model on five public datasets and demonstrate its superiority over benchmark models, achieving better results in terms of RMSE and MAE.

The model used in this MRP shares a similar framework with the one proposed by Wang et al. (2022) (17). Their model employs an attention-based RNN encoder-decoder model for multi-step forecasting of sensory time series. The model captures interrelationships among heterogeneous

parameters, employing spatial and parameter-wise attention layers in the encoders and temporal attention in the decoders for multi-step forecasting. However, there are notable differences in the approach adopted in this MRP versus Wang et al. (2022) (17) . First, the model in this paper focuses on economic data (CPI) instead of sensory data (temperature). Additionally, the incorporation of the teacher forcing technique aids in enhancing information flow and addressing the vanishing gradient problem encountered in the framework. Furthermore, this MRP introduces a normalization layer to enhance model stability and training convergence.

3   Exploratory Data Analysis

The dataset used for this research provides monthly Consumer Price Index (CPI) values for different categories of goods and services. The dataset contains not seasonally adjusted CPI values for all-items, as well as categories such as food, shelter, clothing, transportation, and healthcare. The CPI indices are not seasonally adjusted as exploratory data analysis revealed that there was not much seasonality with the data .

3.1   Data Acquisition

The data for this project was acquired from Statistics Canada (`www150.statcan.gc.ca`). The datasets are open-sourced and compliant with the MRP requirements.

3.2   Data Source & Data Files

The dataset was downloaded as csv files from the Statistics Canada website. The CPI values are reported as indices with the base year set to 2002=100. A subset of the data spanning from January 2003 to the present was extracted for the forecasting process. The data files are posted on GitHub, and the link to the GitHub repository and a sample file are included in Appendix A.

3.3   Data Structure Analysis

The dataset includes fields such as the reference date period (grouped by month), CPI values for different categories of goods and services, and CPI values attributed to energy, goods, and services. Table 1 represents the data fields and their descriptions.

Table 1: Data Fields and Descriptions

Field	Description
Date	The reference date period for which the data is reported. The data is typically grouped by month.
All-Items	The values attributed to the basket of goods used to calculate the overall index. It includes all items, such as food, shelter, clothing, transportation, health care, recreation, and more.
Food	The values attributed to the basket of goods that encompass food items. This category includes food consumed at home and food purchased away from home.
Shelter	The values associated with housing and accommodation costs, such as rent, mortgage payments, and maintenance expenses.
Household operations, furnishings, and equipment	The values attributed to household spending on daily operations, equipment, and furnishings. This category includes expenses like cleaning supplies, appliances, furniture, and other household items.
Clothing and footwear	The values attributed to the cost of clothing and footwear. This category includes clothing items, shoes, and related accessories.
Transportation	The values attributed to the cost of transportation, including vehicle purchases, fuel, vehicle maintenance, public transportation fares, and other transportation-related expenses.
Gasoline	The values associated with the price of gasoline. It reflects the changes in the cost of fuel used for transportation.
Health and personal care	The values attributed to health care and personal care expenses. This category includes medical services, prescription drugs, personal care products, and other related costs.

Field	Description
Recreation, education, and reading	The values attributed to education, recreation, and reading expenses. This category includes costs related to recreational activities, tuition fees, school supplies, books, and other educational materials.
Alcoholic beverages, tobacco products, and cannabis	The values attributed to the cost of alcoholic beverages, tobacco products, and recreational cannabis. This category includes prices of alcoholic beverages, cigarettes, cigars, and other tobacco-related products.
All-Items excluding food and energy	The value attributed to all items except for the prices of food and energy. It provides an indication of overall trends excluding the impact of food and energy costs.
All-Items excluding energy	The value attributed to all items except for the cost of energy.It gives insight into the overall index while excluding the influence of energy prices.
Energy	The value attributed to the cost of energy, including electricity, natural gas, and other energy-related expenses.
Goods	The values attributed to the cost of goods, which includes both durable and non-durable goods. Durable goods are items with a longer lifespan, such as appliances and furniture, while non-durable goods are consumed quickly, such as food and clothing.
Services	The values attributed to the cost of services, including housing services, transportation services, medical services, and other services consumed by households.

3.4 Data Analysis and Statistics

The dataset consists of 244 rows and includes 1 dependent variable (All items category) and 15 independent variables. Summary statistics for each category are presented in Figure 1, providing insights into the distribution, central tendency, and variability of the data.

```
1  #checking summary statistics
2  data.describe().transpose()
```

	count	mean	std	min	25%	50%	75%	max
All-items	244.0	123.229508	13.362860	102.0	112.175	122.70	133.000	156.4
Food	244.0	132.002459	20.085299	100.7	112.600	131.80	144.700	183.1
Shelter	244.0	129.881148	16.165007	102.1	119.500	127.95	140.075	169.9
Household operations	244.0	114.036885	9.665732	100.1	104.000	114.05	122.725	133.7
Clothing and footwear	244.0	94.535246	2.662912	87.9	92.775	94.65	96.225	100.8
Transportation	244.0	128.072541	15.869033	103.7	116.700	127.35	137.125	176.2
Gasoline	244.0	161.768852	33.503064	99.3	142.525	157.40	183.350	299.4
Health and personal care	244.0	117.973361	10.583401	100.5	107.900	118.40	125.600	145.7
Recreation & education	244.0	108.413115	7.160660	99.1	102.075	106.40	114.500	127.1
Alcohol, tobacco & cannabis	244.0	144.764754	21.909026	106.8	126.600	139.60	166.475	189.4
All-items excluding food and energy	244.0	118.772131	11.355626	101.5	109.600	117.00	127.525	146.6
All-items excluding energy	244.0	121.115574	12.842375	101.5	109.975	119.80	130.600	152.7
Energy	244.0	149.719262	24.265764	103.7	134.500	149.80	159.950	242.5
Goods	244.0	115.347951	9.581440	100.9	107.875	114.45	120.200	143.9
Services	244.0	131.046721	17.354589	102.1	116.800	130.20	144.675	168.5

Figure 1: Summary Statistics.

Figure 2, uses a time series plot to illustrate the trends of the various categories. Overall, CPI values have shown a consistent increase month over month, with the exception of clothing and footwear. Notably, categories such as gasoline, energy, and alcohol and cannabis exhibit high volatility, featuring significant peaks and exponential growth since 2003.
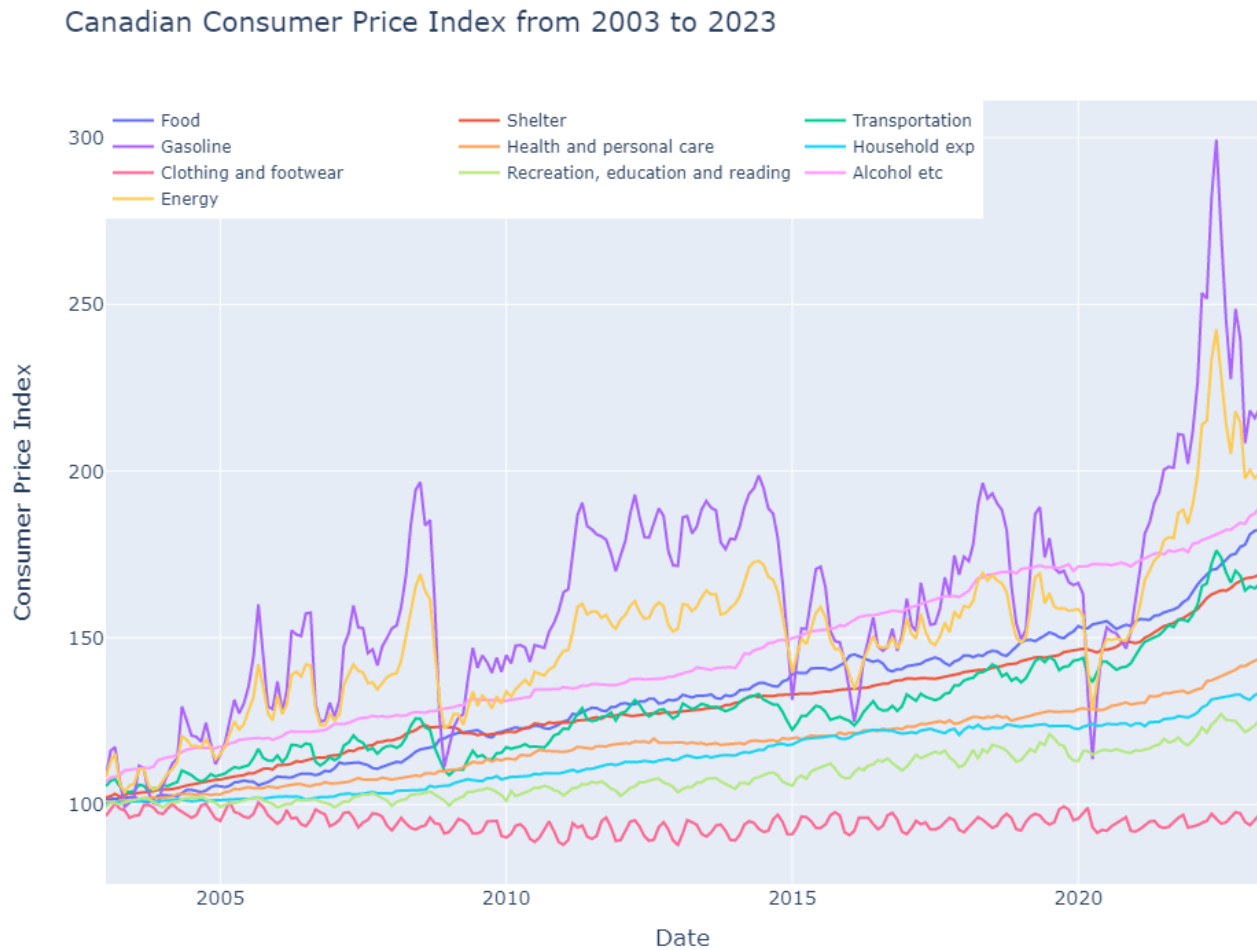

Figure 2: Canadian Consumer Price Index 2003 to 2023

Figure 3, provides an overview of the average Consumer Price Index (CPI) per product category. The data reveals that, on average, CPI values have increased by 23% across all items. Notably, certain categories show even higher average increases. Gasoline, energy, alcohol, food, and shelter have experienced average increases of 60%, 50%, 44%, 31%, and 30%, respectively. These findings indicate that households, on average, allocate a significant portion of their spending—30% or more—towards these specific products. Gasoline, energy, alcohol, food, and shelter play a substantial role in the average household budget, reflecting the impact of rising prices within these categories.
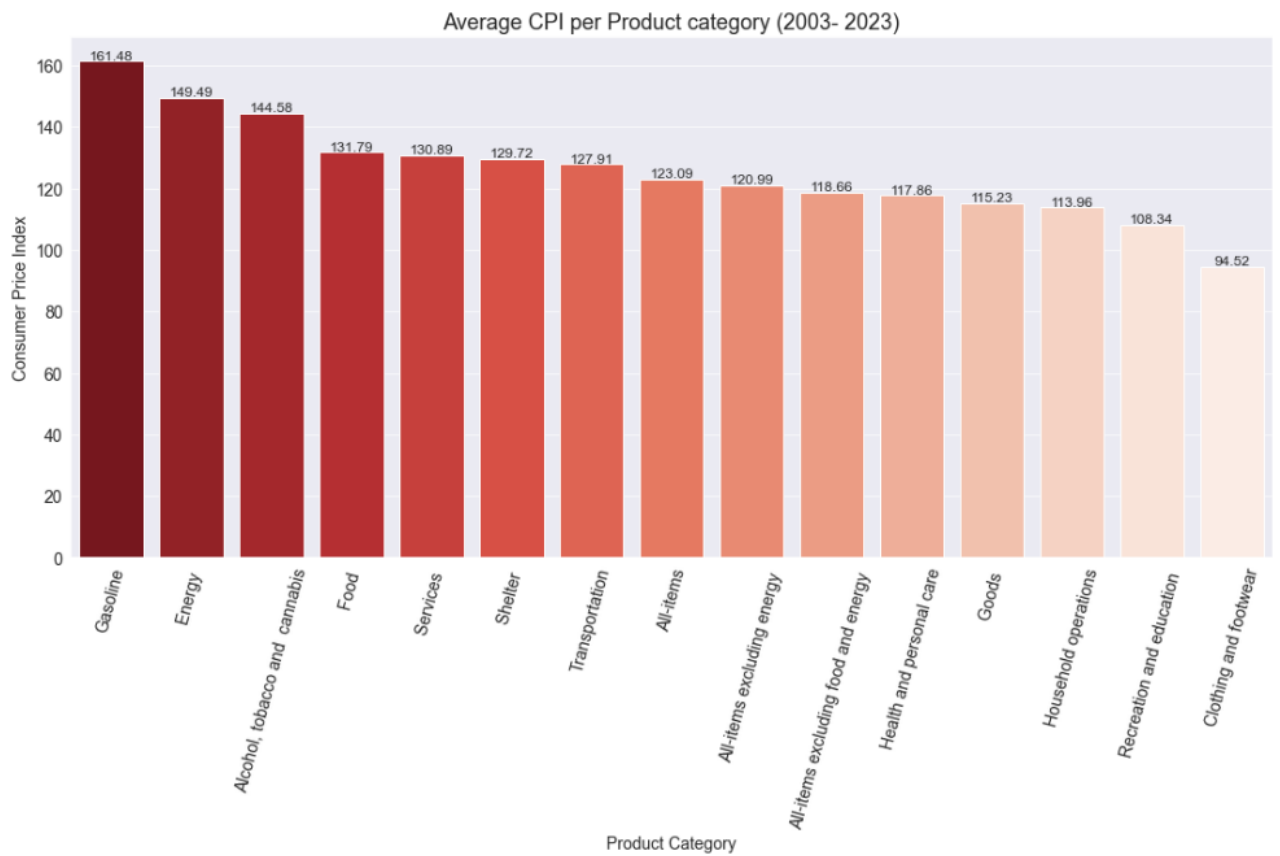

Figure 3: Average CPI by product category

The distribution of the data is illustrated in Figure 4 through box plots. It is evident that most product categories have a lower quartile value greater than 100, except for clothing. This suggests that, compared to 2003, the Consumer Price Index (CPI) has increased for all categories except clothing. The upper quartiles vary across different product categories. Notably, categories such as transportation, gasoline, and energy exhibit extreme outliers. These outliers represent data points that deviate significantly from the overall pattern or distribution of the dataset. They are observations that lie far away from the majority of the data points and can have a disproportionate impact on statistical measures and forecasting results. During the modeling process, it may be necessary to remove these outliers from the training set to prevent them from causing forecasting errors. By removing these extreme values, a higher degree of accuracy and reliability can be ensured in predictions.
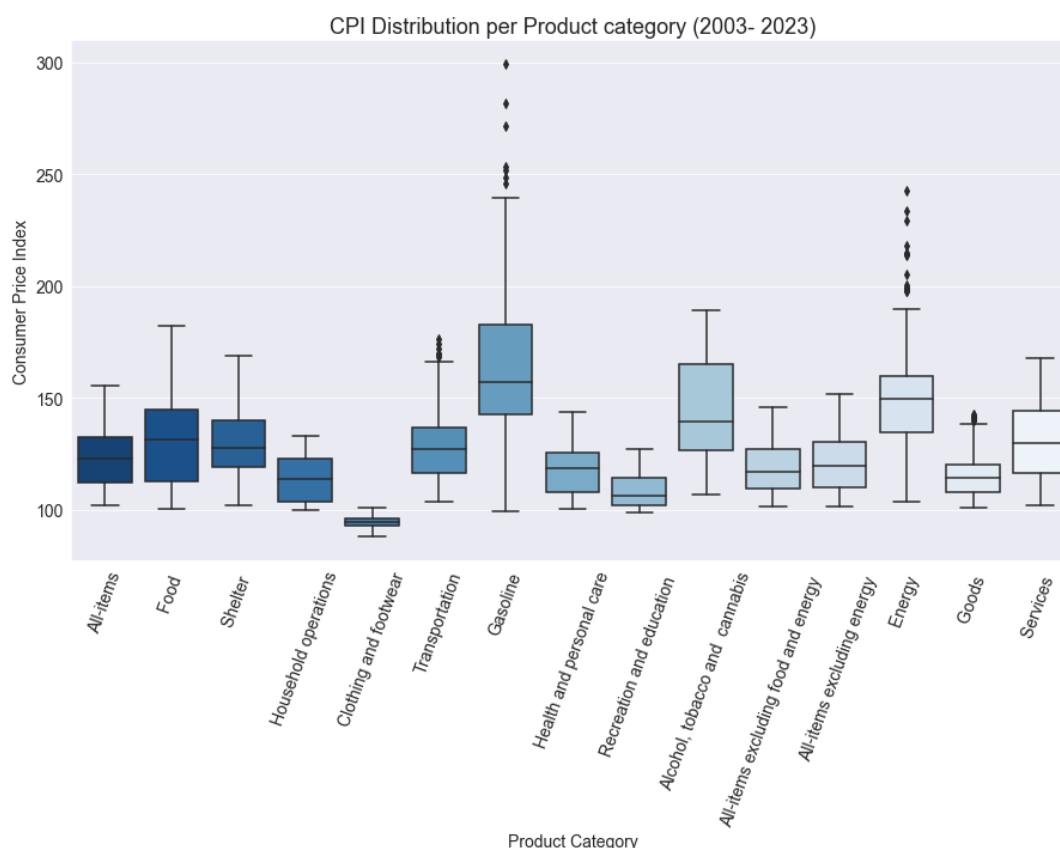

Figure 4: CPI Distribution per Product Category

In Figure 5, the overall Consumer Price Index (All-Item Category) data is shown using a heatmap. The heatmap reveals that there are no discernible patterns or variations in CPI values by month, indicating the absence of monthly seasonality. However, when observing the heatmap over a longer time frame, the colour deepens, particularly in the last two years. This deepening color indicates a consistent growth in CPI compared to the base period of 2002. These findings suggest that while there are no specific months with notably high CPI values indictive of seasonality, the overall trend of CPI has exhibited a steady increase since 2003, with a more pronounced growth in recent years.
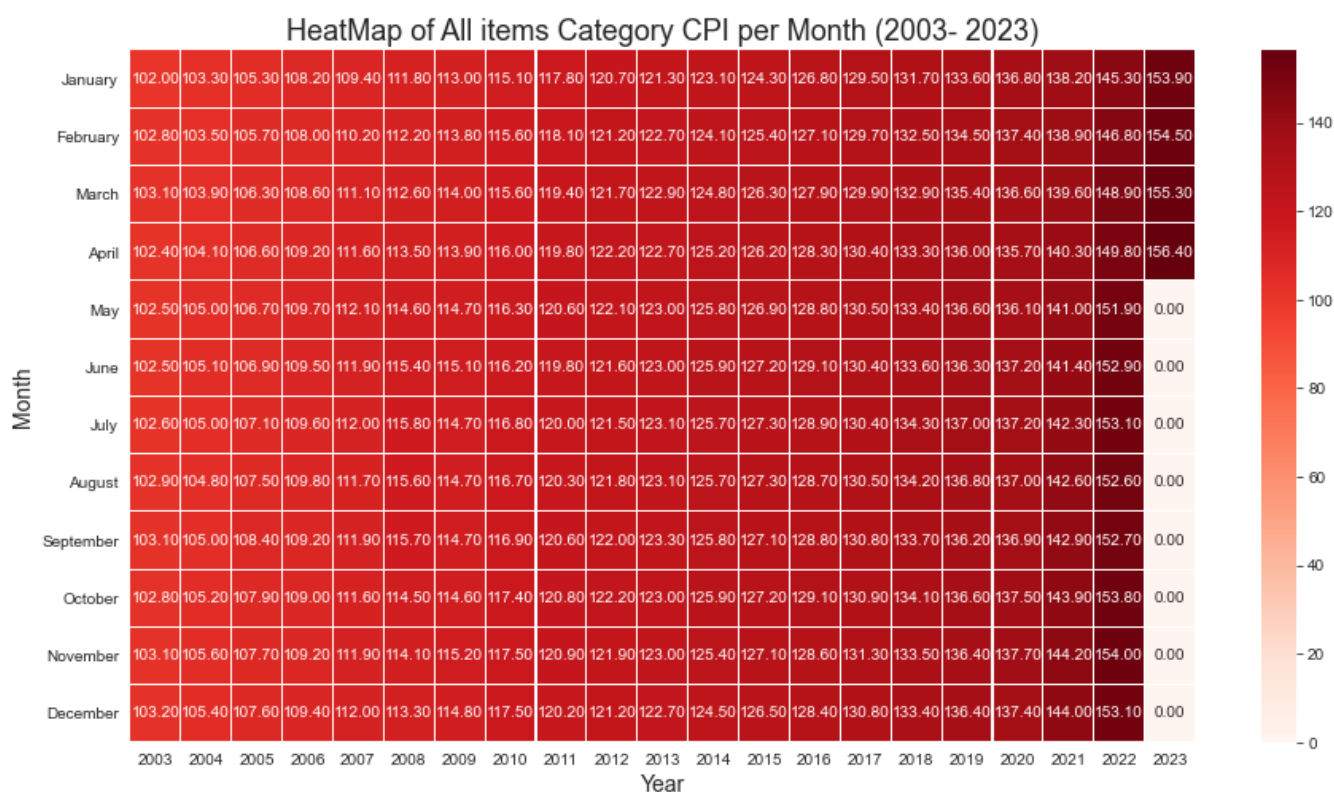
Figure 5: Heat Map of All Item Category

15

The ETS seasonal decomposition in Figure 6 further supports the observation that the data does not exhibit strong seasonality. The decomposition analysis reveals that the seasonal component is relatively weak or negligible.
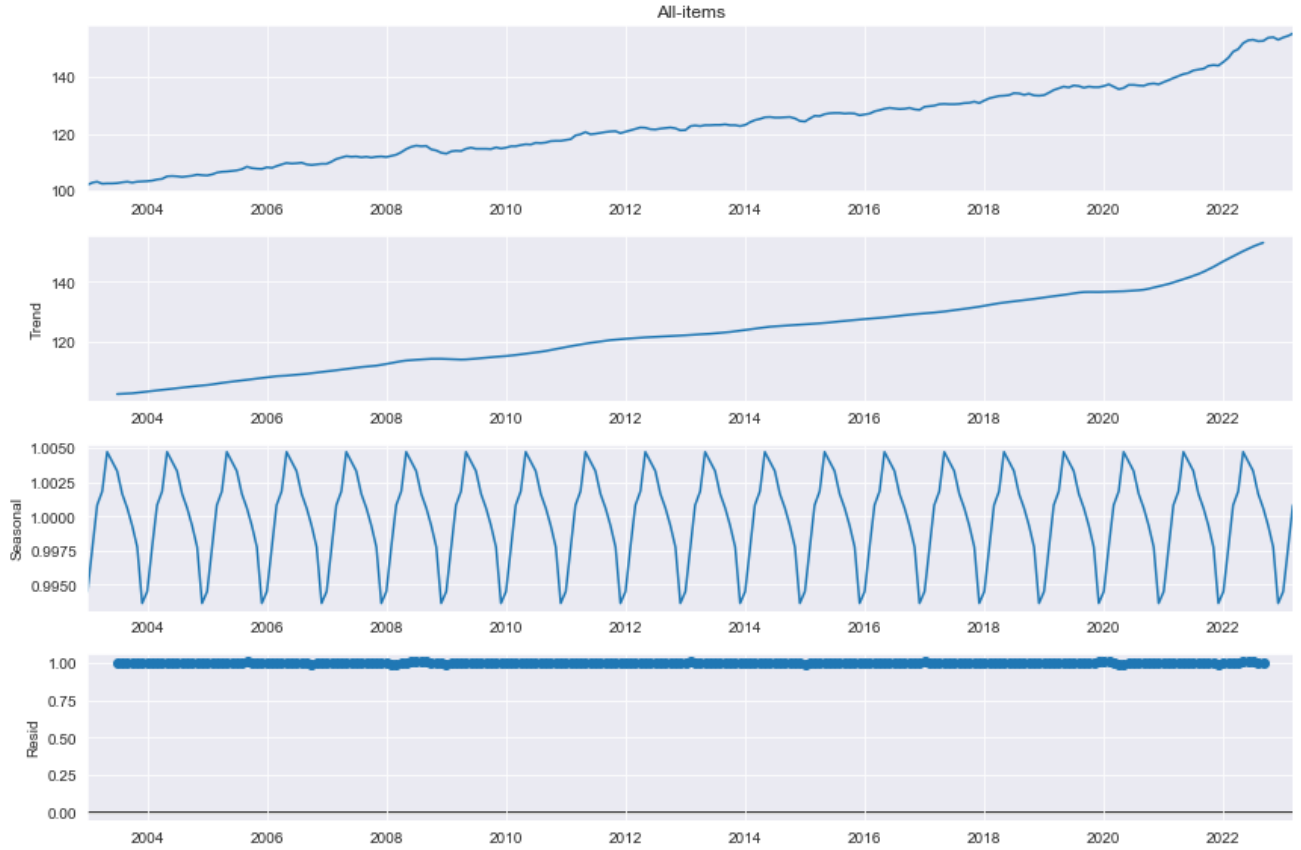
Figure 6: ETS Seasonal Decomposition

Additionally, the Autocorrelation plot in Figure 7 demonstrates that the values in the "All-items" category are non-stationary. The plot illustrates a slow decay or gradual decline in autocorrelation values as the lag increases. This pattern is indicative of non-stationarity in the data. These findings
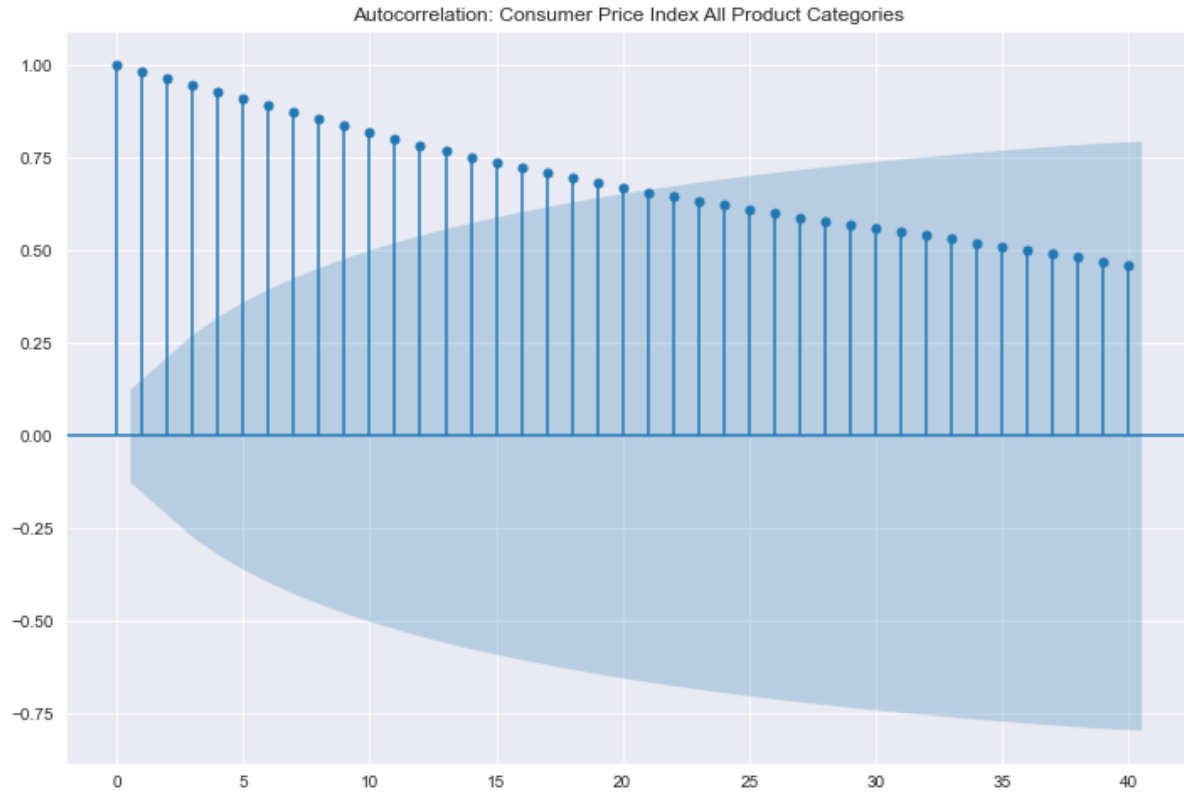
Figure 7: Autocorrelation Overall CPI

suggest that the dataset lacks strong seasonality and exhibits non-stationary behavior, as indicated by both the ETS seasonal decomposition and the Autocorrelation plot.

4   Problem Statement

The problem of interest in this MRP is a multivariate time series problem. A multivariate time series problem involves analyzing a sequence of observations where each observation consists of multiple variables, aiming to capture dependencies and patterns to make predictions or forecasts (Bjerregaard et al., 2021) (21). Simply, the goal is to model future state based on existing historical data. For the CPI data, the anticipation of the future value t+1 is rooted in historical values of both endogenous and exogenous variables within the dataset. Despite the utilization of a deep learning approach in the proposed solution, the data can also be translated into a basic linear regression problem, mathematically formulated as:

$$
\begin{aligned}
\text{CPI}(t) =& \beta_0 + \beta_1 \cdot \text{All-items}(t-1) + \beta_2 \cdot \text{Food}(t-1) + \beta_3 \cdot \text{Shelter}(t-1) \\
& + \beta_4 \cdot \text{Household operations}(t-1) + \beta_5 \cdot \text{Clothing and footwear}(t-1) + \beta_6 \cdot \text{Transportation}(t-1) \\
& + \beta_7 \cdot \text{Gasoline}(t-1) + \beta_8 \cdot \text{Health and personal care}(t-1) \\
& + \beta_9 \cdot \text{Recreation \& education}(t-1) + \beta_{10} \cdot \text{Alcohol, tobacco \& cannabis}(t-1) \\
& + \beta_{11} \cdot \text{All-items excluding food and energy}(t-1) \\
& + \beta_{12} \cdot \text{All-items excluding energy}(t-1) + \beta_{13} \cdot \text{Energy}(t-1) \\
& + \beta_{14} \cdot \text{Goods}(t-1) + \beta_{15} \cdot \text{Services}(t-1) + \varepsilon(t)
\end{aligned}
$$

Where:

- $CPI(t)$ represents the CPI value at time t (the target variable).

- $All\text{-}items(t\text{-}1)$, $Food(t\text{-}1)$, $Shelter(t\text{-}1)$, and so on, represent the CPI values of the corresponding variables at time t-1 (the lagged values of the independent variables).

- $\beta_0, \beta_1, \beta_2, \ldots, \beta_{15}$ are the coefficients (parameters) of the model that need to be estimated.

- $\varepsilon(t)$ represents the error term at time t, which captures the unexplained variation in CPI that is not accounted for by the model.

5   Methodology

The proposed method for the CPI forecasting problem is the Recurrent Neural Network (RNN) Encoder-Decoder model with self attention layers using teacher forcing technique. Before discussing the model, it is essential to provide a brief overview of the fundamental RNN architecture. RNNs are neural networks designed to model sequential data, where each value is influenced by preceding values (Mandic & Chambers, 2001) (18). There are two types of RNNs commonly used: LSTM and GRU. Refer to Figure 8 for a basic depiction of RNN, LSTM, and GRU architectures.
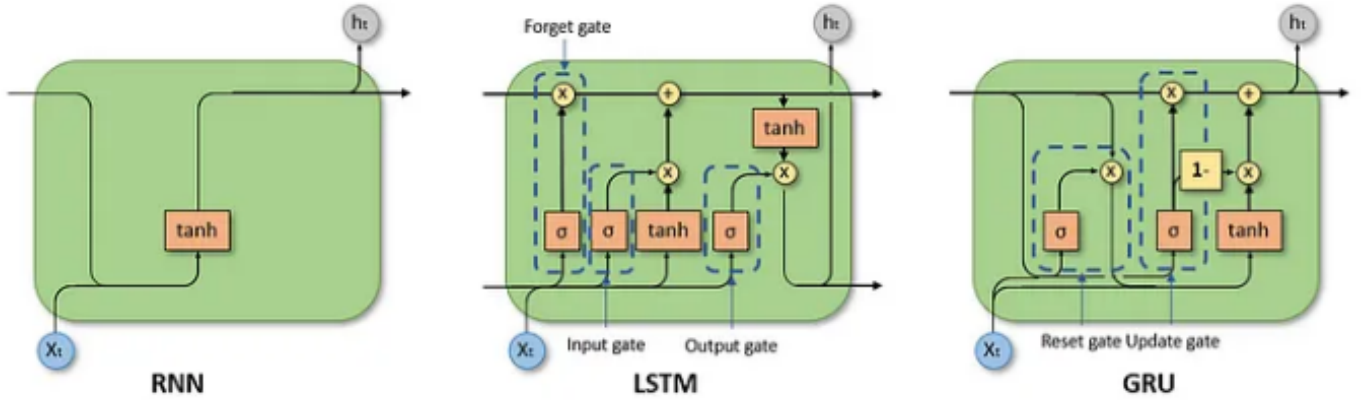
Figure 8: Basic RNN LSTM GRU (adapted from Danker, 2022 (24))

5.1   RNN Attention Encoder Decoder Model

The RNN Encoder-Decoder model is an advanced version of the recurrent neural network architecture first introduced by Cho et al.(2014) (22) . The model's goal is to encode input sequences into a fixed-length vector representation, and then decode them to generate output sequences (Cho et al, 2014) (22). The encoder component captures the input sequence information, while the decoder component generates the corresponding output sequence. The encoder hidden states is

mathematically computed as

$$h_t = f(W_{hh}h_{t-1} + W_{hx}x_t)$$

whereas the decoder hidden states is mathematically compute it as:

$$h_t = f(W_{hh}h_{t-1})$$

In the proposed model the output layer will use a softmax activation function which can also be computed as:

$$y_t = \text{softmax}(W_s h_t)$$

See Appendix B for the description of each variable in the above formulas.

The RNN Encoder-Decoder model has demonstrated significant success in various applications, such as machine translation, speech recognition, and text generation. Researchers have identified that the integration of attention layers can greatly improve this model. By incorporating attention layers, these models are liberated from the limitation of encoding input sequences into fixed-length vectors (Zhang et al.,2014) (23). This advancement allows for more flexible and effective information processing, enhancing the model's ability to capture important dependencies across the input sequence. See Figure 9 for a visual depiction of the RNN Attention Encoder Decoder Architecture.

Although the RNN Encoder-Decoder model has been applied to various forecasting tasks, such as earthquake size prediction by Li et al. (2022) (19) and electricity load forecasting by Xiong et al. (2021) (20), it has not yet been applied to CPI data. The proposed model stands out from existing literature not only due to its application to macroeconomic data but also because of the incorporation of attention layers and the introduction of the teacher forcing technique. The teacher force technique involves using the true target sequence as input for the decoder, rather than using the predicted output from the previous time step (Deng & Pan, 2021) (26). Teacher forcing stabilizes the training process by providing accurate target values as inputs, thereby preventing error propagation and accelerating learning. It ensures more precise and informative inputs for the decoder, allowing the model to converge faster and improve gradient flow. See Figure 10 for basic
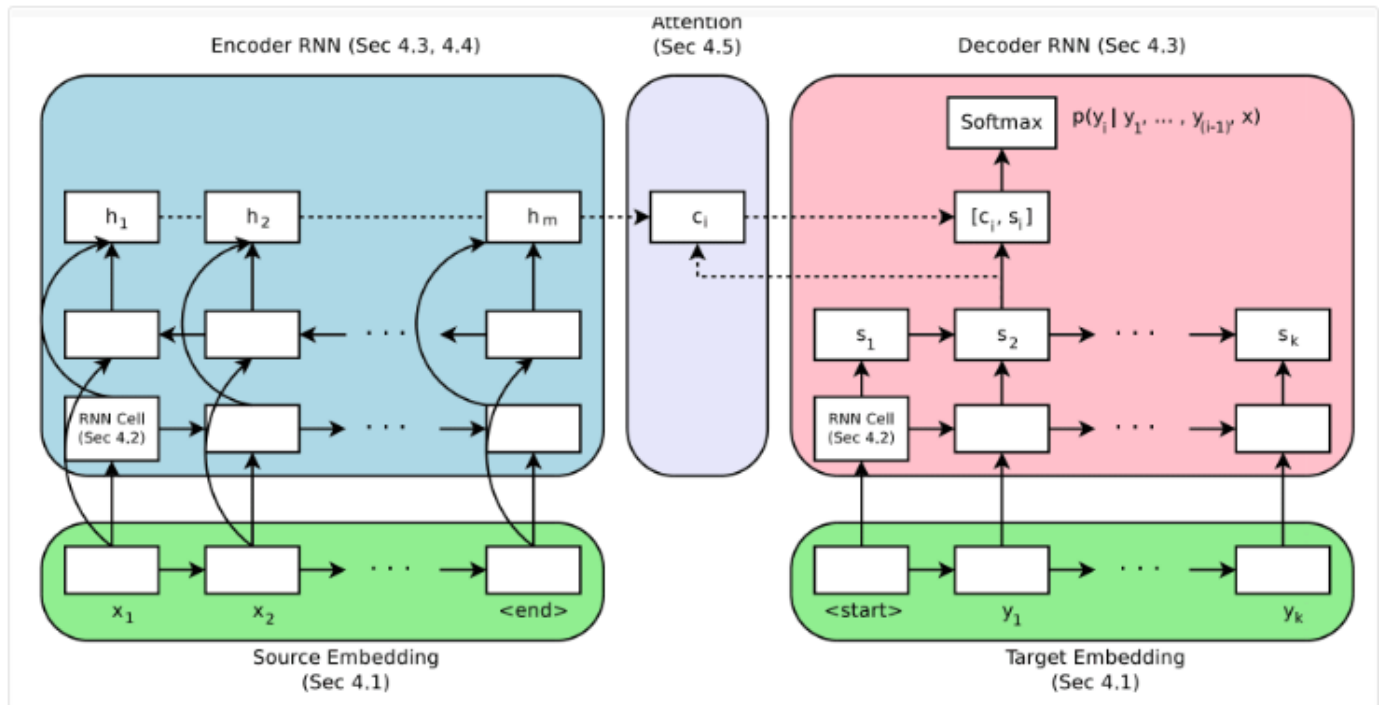
Figure 9: RNN Attention Encoder Decoder Model (adapted from Brownlie, 2018 (25))

architecture of RNN using Teacher Forcing techniques.
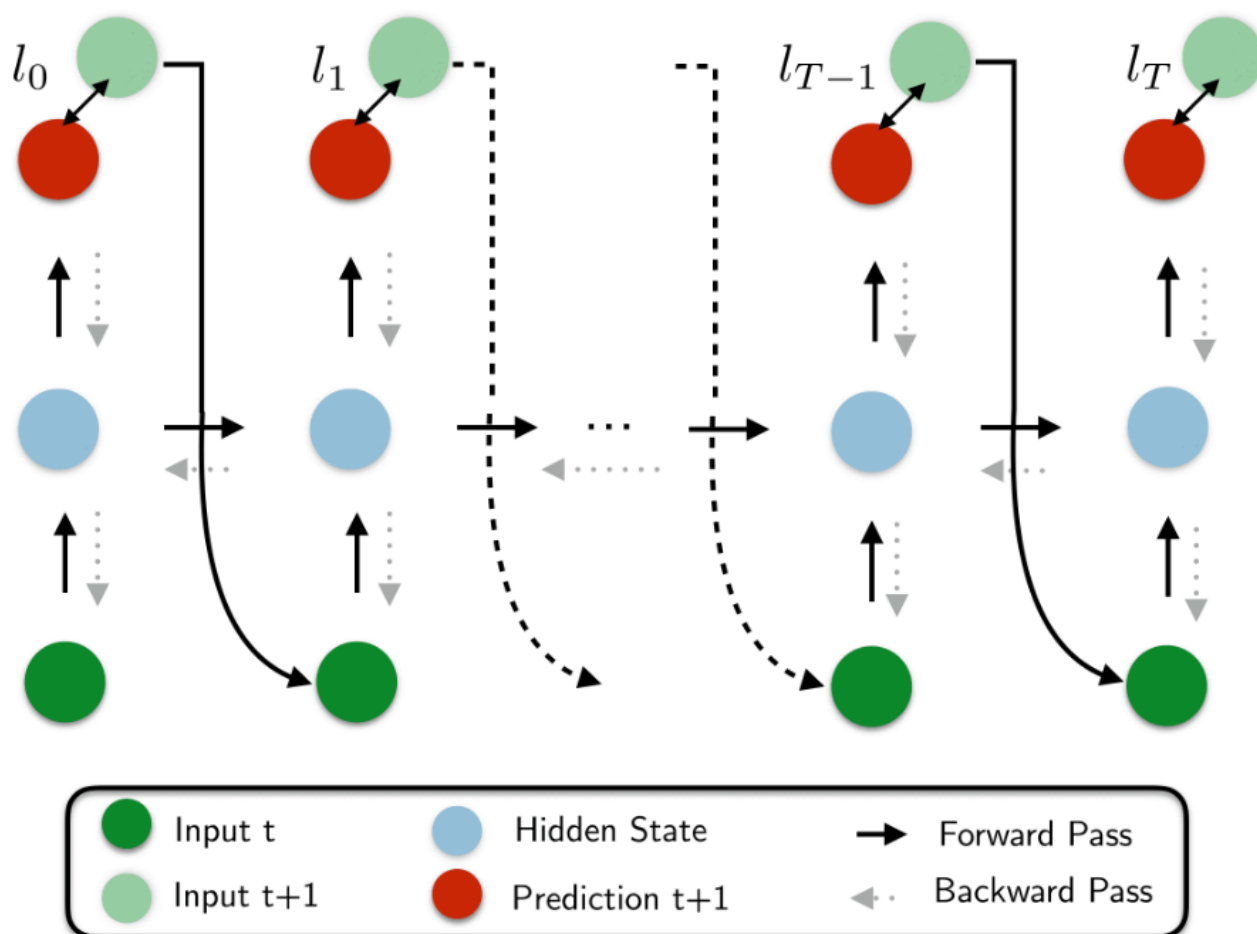
RNN Training with Teacher Forcing

Figure 10: RNN with Teacher Forcing (adapted from Lungu, 2020 (27))

6 Experimentation

6.1 Aim of Study

The aim of this study is to explore the potential of encoder decoder RNN attention architecture in improving the forecasting accuracy of Consumer price index. The model is compared with to baseline models such as Variational Autoregressive models, ARIMA models, Support Vector Regressors, Random Forest Regressors, Artificial Neural Network (ANN) architecture, and Convolutional LSTM architecture to assess its performance.

6.2 Dependent and Independent Variables

The independent variables in this experiment are: Food, Shelter, Household operations, Clothing and footwear category, Transportation, Gasoline, Health and personal care category, Recreation & education, Alcohol, tobacco & cannabis, All-items excluding food and energy, All-items excluding energy, Energy, Goods and Services. The dependent variable is the CPI index value associated with each category.

6.3 Factors and Levels

In this experiment, the factors are the different statistical models with different parameter settings that can validate the model algorithm and its evaluation metric performance.

6.4 Experimental Design

Data Preprocessing

This work utilizes monthly CPI data released by Statistics Canada. The dataset spans from January 2003 to April 2023 and was downloaded as a CSV file from the Stats Can website. The dataset had no missing values, but there were outliers which were removed before experimentation.

Feature Selection and Experimental Design

The data was split into training, validation, and test sets. The proposed model has a training and validation set consisting of the first 232 months, while the test set includes the last 12 months.

Within the training and validation set, the first 80% of the measurements are used for training, and the remaining 20% are used for validation.

Cross Validation

To compare different forecasting models, a validation technique is implemented. Time series cross-validation is used in this project to measure accuracy using R2 score, SMAPE (Symmetric mean absolute percentage error), RMSE (Root Mean Square Error), and MAPE (Mean Absolute Percentage Error) between actual and predicted data

6.5 Experiment Performance and Revisions

A series of experiments were conducted to fine-tune the predicted values and improve the model's R2 values, root mean squared error, mean percentage error, and SMPAE. Details of the different experiments are provided below.

Experiment 1: Proposed Model (Attention RNN Encoder Decoder Model).

Encoder:
The proposed model begins with an encoder layer. The encoder part of the model takes the input sequence (the CPI index values) as input. In the dataset this is a sequence of one-dimensional data. The input sequence is fed into an LSTM (Long Short-Term Memory) layer. The LSTM layer has 180 units and is set to return the sequences as well as the final hidden and cell states. The output of the encoder LSTM layer is stored in encoder_outputs, and the final hidden and cell states are stored in state_h and state_c, respectively. These states capture the information from the input sequence and will be used as the initial states for the decoder.

Decoder:
The decoder part of the model takes the target sequence as input, which is also a sequence of one-dimensional data for the CPI data. During training, the decoder input sequence is the true target output sequence. The input sequence is passed through an LSTM layer with 180 units. The

LSTM layer is set to return the sequences as well as the hidden and cell states. The output of the decoder LSTM layer is stored in decoder outputs, but the hidden and cell states are not used directly in this model. Instead, they are used to initialize the decoder states with the information captured by the encoder. This allows the decoder to generate an output sequence based on the input sequence.

Attention Mechanism:

The attention mechanism is a crucial component of this model. It helps the decoder focus on relevant parts of the input sequence during the generation of each output. In this implementation, the attention mechanism is implemented using the dot product attention. First, the dot product between the decoder outputs (decoder_outputs) and the encoder outputs (encoder_outputs) is computed using the Dot layer with axes=[2, 2]. This operation calculates the similarity between the decoder outputs and each encoder output. Next, an activation function (softmax) is applied to normalize the similarity scores across the encoder outputs. This step allows the model to assign higher weights to more relevant parts of the input sequence. The normalized attention weights are stored in the attention variable. The attention weights are then used to compute a weighted sum of the encoder outputs, giving more importance to the parts of the input sequence that the model pays more attention to. This is done using the Dot layer with axes=[2, 1] and stored in the context variable. Finally, the attention context vector and the decoder outputs are concatenated along the last axis using the Concatenate layer.

Output and Training:

The concatenated context vector and decoder output are passed through a TimeDistributed dense layer with linear activation, which produces the final output sequence. During training, instead of using the previous decoder output as input for the next step, the true target output sequence is fed into the decoder. This incorporates the teacher forcing technique. The model is compiled with the Adam optimizer and mean squared error (MSE) loss function. It is then trained using the training data and the corresponding target output for 500 epochs with a batch size of 4. After training, the model is evaluated on the test set. The test input sequences are fed into the model to generate predictions. The predictions are then compared with the target output sequence. See Figure 11 for visualization of model architecture.
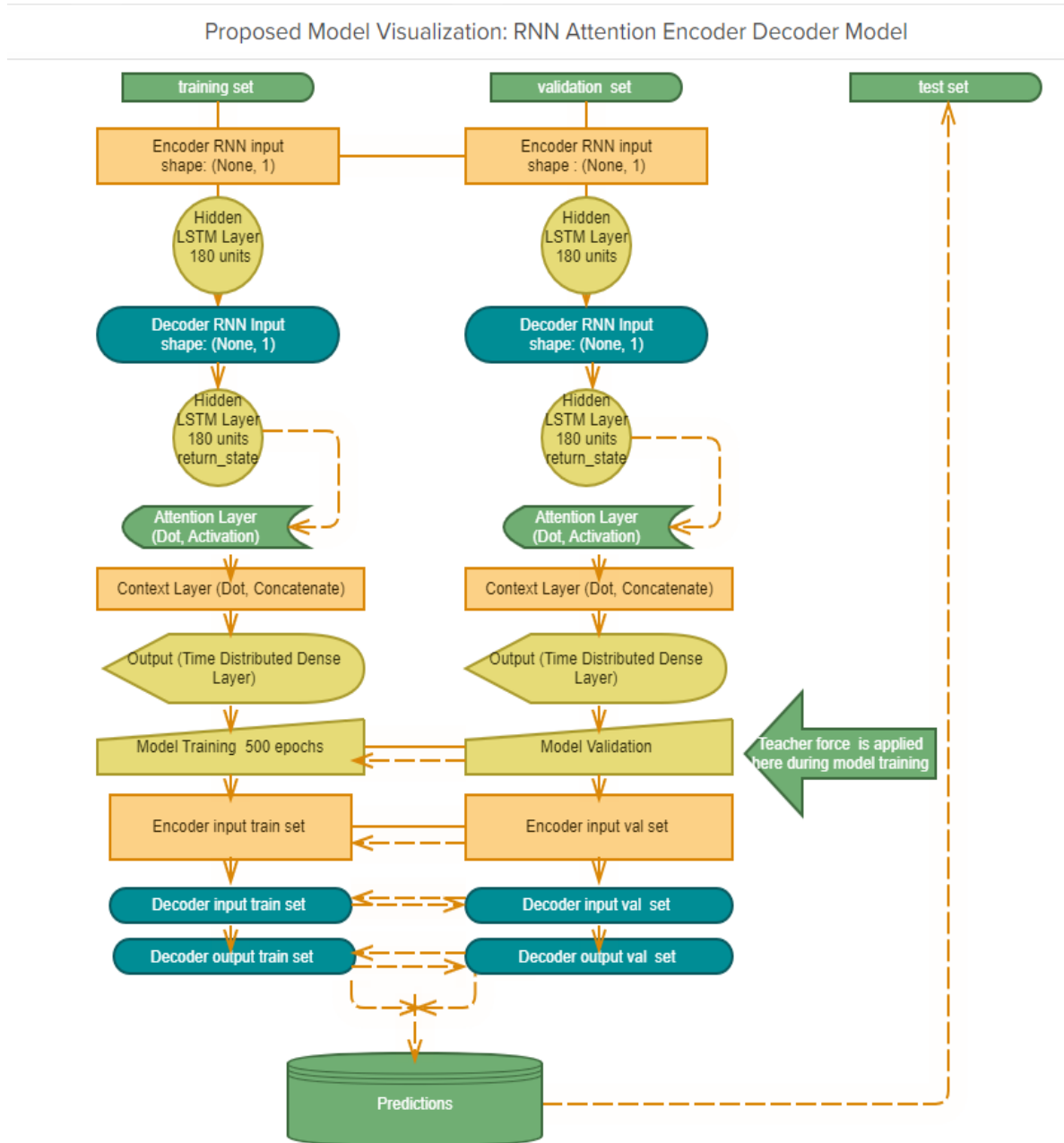
Figure 11: RNN Attention Encoder Decoder Model with Teacher Force Technique

Experiment 2: Baseline Models: Traditional Statistical Techniques

In this section, a comparison is drawn between the proposed model and the ARIMA and VAR models. Dondong (2010) (4) advocated for the use of ARIMA and VAR models as effective models for CPI forecasting.

Variational Autoencoder Model:

For the Variational Autoencoder baseline model, the dataset was divided into a training set (238 months) and a test set (6 months). The VAR model was trained on the training set using the VAR class, and the optimal lag order was determined using the Akaike Information Criterion (AIC). Autocorrelation and partial autocorrelation plots were used to analyze residuals, and CPI was forecasted for the test set.

ARIMA Model:

For the ARIMA model, the dataset was also divided into a training set (238 months) and a test set (6 months). For each CPI indicator the auto arima feature was implemented to find the appropriate selected order. Stationarity was checked using the ADF test, and ARIMA models with the specified order were fitted for each component of the CPI data. Autocorrelation and partial autocorrelation plots were generated to analyze residuals, and forecasting was performed on the test set.

Experiment 3: Baseline Models: Supervised Learning Techniques

In this section, the implementation of the Lasso regressor, Support Vector Regressor (SVR), and Random Forest Regressor models is discussed, serving as benchmarks against the proposed architecture.

Lasso Regressor:

In the Lasso regressor experiment, the dataset was split into a training set (238 months) and a test set (6 months). Sequential values were used for training, and a Lasso regressor model was created with an alpha value of 0.1. Multiple Lasso regressors were combined using Multi Output Regressor, which integrated their predictions for the final forecast. The Multi Output Regressor was trained on the scaled training data and used to predict values on the scaled testing data.

Support Vector Regressor:

For the Support Vector Regressor experiment, the dataset was divided into a training set (238 months) and a test set (6 months). The Support Vector Regressor (SVR) method was applied separately to each CPI component. The training data, consisting of sequential values, was reshaped into a 2D array to fit the SVR model. Subsequently, predictions were generated for the test data using the trained SVR models. The forecasts from each SVR model were combined. Finally, the forecast was used to predict values for the test set.

Random Forest Regressor:

In the Random Forest Regressor experiment, the dataset was split into a training set and a test set. Each CPI component was trained individually using the first 97% of the data for training and the last 3% of the data for testing. This was done to get a 238 training set and 6 month test set. Sequential integers were used to represent the training data. A Random Forest Regressor model was created and used to predict values on the test set. The model's residuals, representing the difference between actual and predicted values, were analyzed through autocorrelation and partial autocorrelation plots.

Experiment 4: Baseline Models: Deep Learning Techniques

In this section, experiments are conducted using the Artificial Neural Network (ANN) and Convolutional LSTM models as baselines to compare to the proposed architecture.

Artificial Neural Network (ANN) Model:

In the implementation of the Artificial Neural Network (ANN) experiment, the dataset was divided into training and test sets. The training set contained 97% of the CPI indices, while the remaining 3% formed the test set. The ANN model was constructed using TensorFlow's Sequential API, consisting of three dense layers with 180 units each. All layers, including the output layer, utilized the ReLU activation function. The model was compiled with the Adam optimizer and the mean squared error (MSE) loss function. Training was performed for 500 epochs with a batch size of 2. Subsequently, the trained model was used to make predictions on the test set.

Convolutional LSTM Model:

For this implementation, the data was split into training and testing sets. The ConvLSTM model

was constructed using TensorFlow's Sequential API. It consisted of a ConvLSTM2D layer with 120 units, followed by a Flatten layer, a fully connected Dense layer with 120 units, and a final Dense layer with a number of units equal to the target columns. The model was compiled with the Adam optimizer and mean squared error loss. Training was conducted using the fit function for 500 epochs with a batch size of 2. After training, the model was evaluated on the test data to compute the loss value, and predictions were made on the test data using the trained model.

7   Results

7.1   Model Evaluation

The results of the proposed and baseline models were evaluated using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), SMAPE (Symmetric Mean Absolute Percentage Error) and R-squared (R2).

Root Mean Squared Error (RMSE):

The RMSE mathematically formulated as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$

Is used provide a measure of the model's accuracy by quantifying the average squared differences between the predicted values and the actual values . The RMSE is preferred evaluation metric because it is sensitive to outliers making it a useful metric for identifying and assessing the impact of extreme predictions.

Mean Absolute Error (MAE):

The MAE mathematically formulated as

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$$

calculates the average absolute difference between the predicted values and the actual values. The MAE is less sensitive to outliers and hence makes it a better choice when the data contains extreme outliers that could significantly impact the prediction results.

Symmetric Mean Absolute Percentage Error (SMAPE):

The SMAPE mathematically formulated as

$$\text{SMAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \times 100$$

is also used to assess the relative accuracy of a model's predictions. It calculates the percentage difference between the predicted values and the actual values in a symmetric manner. It is a preferred evaluation metric because both overestimation and underestimation errors are treated

equally.

R-squared (R2):

Finally, the R2 mathematically formulated as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

measures the proportion of the variance in the dependent variable (in our case CPI indices) that is predictable from the independent variables (CPI predictions). It is a preferred metric due to its goodness of fit nature as well as interpretability. It provides a measure of how well the model fits the data. A higher R2 value indicates that a larger proportion of the variance in the data is accounted for by the model. Again, R2 has a straightforward interpretation. A value of 1 indicates that the model perfectly fits the data, while a value of 0 means that the model provides no improvement over a simple mean-based prediction.

7.2   Model Results

Table 2 provides an overview of the overall results/performance on each experimental model.

| Model	RMSE	MAE	SMAPE	R2
RNN Attention Encoder Decoder Model	**0.1585**	**0.0968**	**0.0762**	**0.9997**
VAR	6.4255	5.3876	0.0309	0.9107
ARIMA	10.4289	4.8552	2.4175	−2.3527
Lasso Regressor	0.4049	0.2853	0.1654	0.9315
Support Vector Regressor	0.2923	0.2063	0.1353	0.9954
Random Forest Regressor	1.5490	0.7897	0.1140	0.9953
Artificial Neural Network	7.4081	2.1677	0.6317	0.9631
Convolutional LSTM	13.9133	8.8546	3.3672	0.3688

Table 2: Performance Comparison Between Proposed vs Baseline Methods

From **Table 2**, we observe that the RNN Attention Encoder Decoder Model achieved the lowest RMSE, MAE, and SMAPE values, indicating smaller prediction errors and more balanced performance. Additionally, the R2 score of 0.99 demonstrates a high level of goodness of fit, indicating that around 99% of the variance in the CPI data can be accurately explained by the model's predictions. In contrast, the other models showed higher errors and lower R2 scores, indicating weaker predictive capabilities.

Below the performance of the proposed RNN Attention Encoder Decoder Model is compared to the baseline models at a 6-time step interval using graphical display. In Figure 8, the RNN Attention Encoder Decoder Model consistently predicts more accurately and closer to the actual values compared to other baseline models. The deviation between the predicted value and the ground-truth value for the proposed model is not too large, while more obvious deviations are observed in the VAR, Lasso Regressor, RFR, SVR, and Convolutional LSTM Models. For instance, when examining the predictions for April 2023, the actual CPI value is 156.4. The proposed model predicts a CPI value of 156.5, whereas the baseline models predict 159.2 using VAR, 156.8 with ARIMA, 155 with Lasso Regressor, 155.5 with Random Forest Regressor, 158.3 with Support Vector Regressor, 156.3 with the ANN model, and 157.2 with the Convolutional LSTM model. This exemplifies that the proposed model predicts much closer to the actual values.
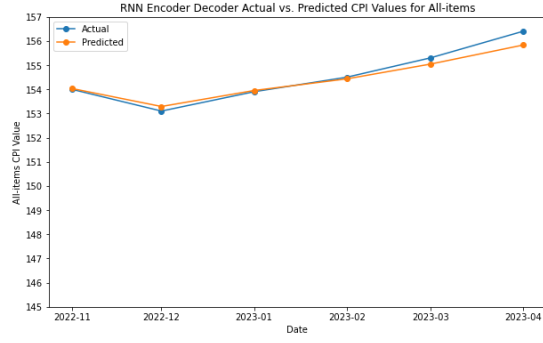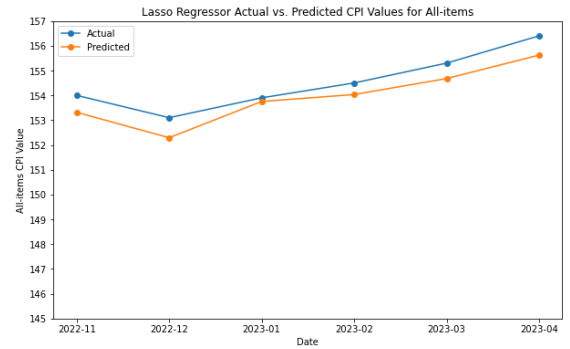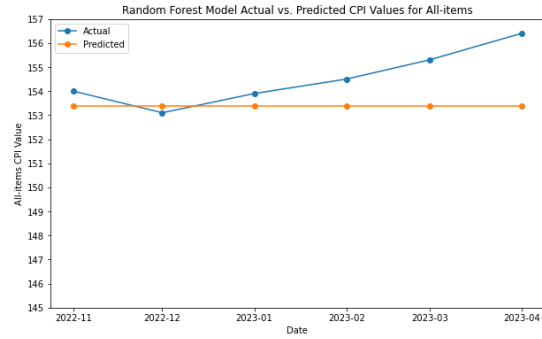
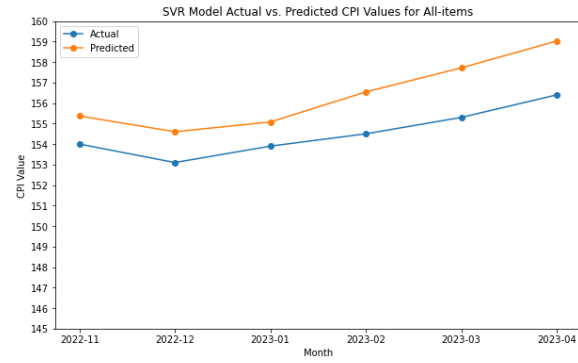(a) RNN Attention Encoder Decoder Model
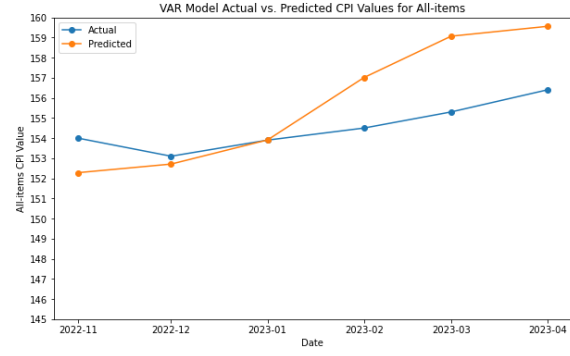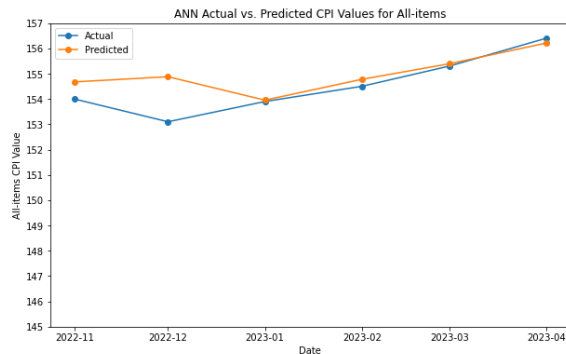
(b) VAR Model
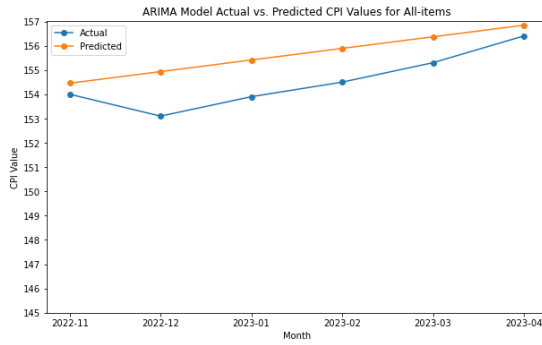
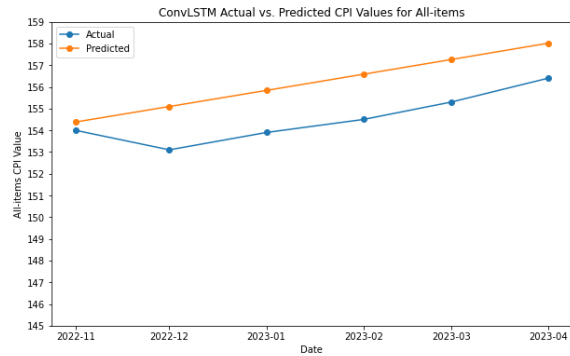(c) ARIMA Model

(d) Lasso Regressor Model

(e) Random Forest Regressor Model

(f) Support Vector Regressor Model

(g) ANN Model

(h) Convolutional LSTM Model

33

Figure 12: Model Experiment with 6 time steps for Actual vs Prediction

In **Table 3** we see performance for RNN Attention Encoder Decoder Moder for each CPI Indicator

CPI Indicator	RMSE	MAE	SMAPE	R2 Score
All-items	0.437 936	0.197 009	0.142 85	0.999 038
Food	0.161 524	0.138 011	0.101 987	0.999 941
Shelter	0.163 271	0.140 608	0.109 276	0.999 905
Household operations	0.155 238	0.137 616	0.123 686	0.999 766
Clothing and footwear	0.153 369	0.103 751	0.111 492	0.996 25
Transportation	0.186 686	0.143 888	0.109 012	0.999 884
Gasoline	1.076 951	0.293 529	0.135 658	0.999 174
Health and personal care	0.149 622	0.126 032	0.107 032	0.999 82
Recreation & education	0.111 623	0.102 69	0.094 262	0.999 803
Alcohol, tobacco & cannabis	0.131 05	0.111 033	0.077 413	0.999 967
All-items excluding food and energy	0.142 433	0.130 271	0.111 262	0.999 859
All-items excluding energy	0.118 856	0.106 983	0.088 302	0.999 923
Energy	0.184 13	0.154 158	0.098 693	0.999 954
Goods	0.137 075	0.126 854	0.110 661	0.999 824
Services	0.133 564	0.112 557	0.085 544	0.999 946

Table 3: RNN Attention Encoder Decoder Model Evaluation Results

From the above, the model exhibits exceptional performance across various CPI indicators. Food, Shelter, Household operations, and Transportation, predictions demonstrated remarkable accuracy, with R2 scores close to 1. The model's impressive performance extends to various other CPI components, including Health and personal care, Recreation & education, and Alcohol, tobacco & cannabis. In each case, the model achieved low prediction errors and high R2 scores, indicating its reliable forecasting capabilities across diverse sectors of the economy. The RNN Attention Encoder Decoder Model also proved valuable for forecasting overall inflation trends. When considering All-items, All-items excluding food and energy, and All-items excluding energy, the model produced precise predictions, as reflected by their R2 scores approaching 1. Although the Gasoline CPI category exhibited slightly higher errors in the RMSE, the model's R2 score of 0.999 still suggests that it captures a substantial proportion of the variance in this indicator. Overall, the predictions

on the various indicators were good.

7.3   Model Forecast

Table 4 presents the 6-month forecast derived from the proposed model for the different CPI indicators.

CPI Index	May/23	Jun/23	Jul/23	Aug/23	Sep/23	Oct/23
All-items	148.89	159.74	162.93	169.06	163.70	163.99
Food	173.16	186.22	189.28	190.04	190.25	190.34
Shelter	165.42	175.19	177.76	178.46	178.74	178.88
Household operations	129.76	140.08	142.68	143.49	143.84	144.03
Clothing & footwear	95.60	96.80	97.50	97.90	96.20	98.20
Transportation	161.49	170.91	174.30	175.82	176.71	177.28
Gasoline	216.31	253.18	261.73	263.78	264.34	264.58
Health and personal care	141.14	152.70	156.26	157.14	157.38	157.49
Recreation, education and reading	123.70	136.76	140.07	140.99	141.31	141.49
Alcohol, tobacco & cannabis	184.59	199.54	204.80	206.08	206.45	206.60
All-items excluding food & energy	139.18	151.90	155.37	156.23	156.54	156.70
All-items excluding energy	150.13	162.29	165.52	166.33	166.56	166.67
Energy	198.23	235.26	245.23	247.66	248.18	248.35
Goods	135.63	153.57	156.92	157.80	158.11	158.27
Services	160.13	169.22	172.57	173.71	174.26	174.59

Table 4: CPI Forecasted Values for May 2023, to October 2023

The predictions indicate that CPI values for all indices, except Clothing and Footwear, are anticipated to rise in the upcoming two quarters, signaling the potential presence of inflationary pressures. In response, monetary authorities, such as the Bank of Canada, may consider implementing a tighter monetary policy by raising interest rates. This measure can help curb inflation by reducing borrowing and spending, potentially stabilizing prices. Additionally, sectors like Gasoline and Energy are projected to experience significant price increases, warranting targeted policies to address sector-specific inflationary pressures, such as promoting energy efficiency or providing subsidies to mitigate consumer impact.

The forecasted rise in CPI for Food and other essential items may lead to increased living costs for

households. Fiscal authorities can consider implementing measures to support vulnerable populations, such as food assistance programs, or direct cash transfers.

While the forecasted trends provide valuable directional cues for policy decisions, it's important to acknowledge that actual outcomes may be subject to uncertainties. Policymakers can use these projections as a basis for formulating flexible and adaptive policies that respond to evolving economic conditions.

8   Conclusion

In this paper, an Attention-based RNN Encoder-Decoder architecture with the Teacher Forcing technique was leveraged to provide a more accurate forecast for inflation indicators such as CPI. The model was designed to bridge the gap between capturing long-term dependencies in CPI data and effectively handle the non-linearity commonly present in such datasets. Additionally, the application of the Teacher Forcing technique served to stabilize and accelerate the model training process while mitigating issues related to error propagation.

The results of the research demonstrated the outstanding performance of the RNN Encoder-Decoder Attention Model, surpassing both traditional approaches and various supervised and deep learning techniques. The model achieved an impressively low RMSE of 0.15, indicating its ability to generate predictions close to the actual CPI values. Furthermore, the high R2 value of 0.99 illustrated the model's superior predictive power, explaining almost all of the variance in the CPI data.

From a policy perspective, the model's outcomes signal an anticipated growth in the Consumer Price Index (CPI), with potential repercussions on inflation rates across Canada. Hence, emphasizing the need for suitable macroeconomic policies.

While this research has achieved promising results, it is essential to acknowledge its limitations. For example interpertability with attention based decoder encoder mechanisms can be challenging that is understanding exactly how the model makes its decisions and which input elements are most influential in the predictions might not be straightforward. Hence it is very difficult to do any kind of backward engineering.

Regardless of its limitations the findings of this research make a significant contribution to the field of economic forecasting and underscore the potential of employing advanced deep learning techniques for CPI prediction. Moreover, this study opens avenues for future research on economic forecasting, including exploring other transformer methods, attention mechanisms, and hyperparameter tuning techniques such as Scheduled Sampling, which shares similarities with Teacher Forcing. As the world continues to address the current inflation crises, the knowledge gained from this study can contribute to making informed decisions and developing effective policies to ensure economic stability and growth.

Appendix A

GitHub Repository

https://github.com/maamecobbinah/Masters-Research-Project-

Data Files

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000401

CPI Reference Paper

https://www150.statcan.gc.ca/n1/en/pub/62-553-x/62-553-x2023001-eng.pdf?st=V_SAU7W0

Additional Resources on Canadian CPI

https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2301#a2

Appendix B

Description of Encoder Decoder mathematical formula

Encoder layer hidden state is calculated as:

$$h_t = f(W_{hh}h_{t-1} + W_{hx}x_t)$$

Where:

- h_t is the hidden state of the encoder at time t.

- W_{hh} is the weight matrix for the hidden state to hidden state transformation.

- W_{hx} is the weight matrix for the input to hidden state transformation.

- h_{t-1} is the previous hidden state of the encoder at time $t-1$.

- x_t is the input at time t.

- f is the activation function.

Decoder layer hidden state is calculated as:

$$h_t = f(W_{hh}h_{t-1})$$

Where:

- h_t is the hidden state of the decoder at time t.

- W_{hh} is the weight matrix for the hidden state to hidden state transformation.

- h_{t-1} is the previous hidden state of the decoder at time $t-1$.

- f is the activation function.

Softmax activation function is calculated as :

$$y_t = \text{softmax}(W_s h_t)$$

Where:

- y_t is the output at time step t.

- softmax is the softmax activation function that normalizes the logits into a probability distribution.

- W_s is the weight matrix for the transformation from the hidden state h_t to the output y_t.

References

[1] Nyoni, T. (2019). ARIMA modeling and forecasting of Consumer Price Index (CPI) in Germany. Federal Reserve Bank of St Louis.

[2] Statistics Canada. (2023). Consumer Price Index Portal. Retrieved from `https://www.statcan.gc.ca/en/subjects-start/prices_and_price_indexes/consumer_price_indexes`

[3] Atkeson, A., & Ohanian, L. E. (2001). Are Phillips Curves Useful for Forecasting Inflation? Federal Reserve Bank of Minneapolis.

[4] Dongdong, W. (2010). The Consumer Price Index Forecast Based on ARIMA Model. 1, 307–310.

[5] Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. Journal of Business & Economic Statistics, 39(1), 98–119.

[6] Wang, Y., Wang, B., & Zhang, X. (2012). A New Application of the Support Vector Regression on the Construction of Financial Conditions Index to CPI Prediction. Procedia Computer Science, 9, 1263–1272.

[7] Liu, J., Ye, J., & E, J. (2023). A multi-scale forecasting model for CPI based on independent component analysis and non-linear autoregressive neural network. Physica A, 609, 128369.

[8] Bhardwaj, V., Bhavsar, P., & Patnaik, D. (2022). Forecasting GDP per capita of OECD countries using machine learning and deep learning models. 1–6.

[9] Nakamura, E. (2005). Inflation forecasting using a neural network. Economics Letters, 86(3), 373–378.

[10] Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., & Koenigstein, N. (2022). Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks. International Journal of Forecasting.

[11] Theoharidis, A. F., Guillén, D. A., & Lopes, H. (2023). Deep learning models for inflation forecasting. Applied Stochastic Models in Business and Industry.

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need.

[13] Abbasimehr, H., & Paki, R. (2022). Improving time series forecasting using LSTM and attention models. Journal of Ambient Intelligence and Humanized Computing, 13(1), 673–691.

[14] Chen, F. (2022). Deep Neural Network Model Forecasting for Financial and Economic Market. *Journal of Mathematics (Hidawi)*, 2022, 1–10.

[15] Li, H., Shen, Y., & Zhu, Y. (2018). Stock Price Prediction Using Attention-based Multi-Input LSTM. Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

[16] Du, S., Li, T., Yang, Y., & Horng, S.-J. (2020). Multivariate time series forecasting via attention-based encoder–decoder framework. Neurocomputing (Amsterdam), 388, 269–279.

[17] Wang, Y., Li, T., Lu, W., & Cao, Q. (2022). Attention-inspired RNN Encoder-Decoder for Sensory Time Series Forecasting. Procedia Computer Science, 209, 103–111.

[18] Mandic, D., & Chambers, J. (2001). Recurrent neural networks for prediction: learning algorithms, architectures and stability. Wiley.

[19] Li, T., Pan, Y., Tong, K., Ventura, C. E., & de Silva, C. W. (2022). Attention-Based Sequence-to-Sequence Learning for Online Structural Response Forecasting Under Seismic Excitation. IEEE Transactions on Systems, Man, and Cybernetics. Systems, 52(4), 2184–2200.

[20] Xiong, J., Zhou, P., Chen, A., & Zhang, Y. (2021). Attention-based Neural Load Forecasting: A Dynamic Feature Selection Approach. 01–05.

[21] Bjerregård, M. B., Møller, J. K., & Madsen, H. (2021). An introduction to multivariate probabilistic forecast evaluation. Energy and AI, 4, 100058.

[22] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078ces

[23] Zhang, D., Fang, Z., Cao, Y., Liu, Y., Chen, X., & Tan, J. (2018). Attention-based RNN model for joint extraction of intent and word slot based on a tagging strategy. In Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27 (pp. 178-188). Springer International Publishing.

[24] Dancker, J. (2022). A Brief Introduction to Recurrent Neural Networks: An introduction to RNN, LSTM, and GRU and their implementation. Retrieved from https://towardsdatascience.com/a-brief-introduction-to-recurrent-neural-networks-638f64a61ff4

[25] Brownlee, J. (2018). How to Configure an Encoder-Decoder Model for Neural Machine Translation. Retrieved from https://machinelearningmastery.com/configure-encoder-decoder-model-neural-machine-translation/

[26] Deng, L., & Pan, Y. (2021). Data-driven proxy model for waterflood performance prediction and optimization using Echo State Network with Teacher Forcing in mature fields. Journal of Petroleum Science & Engineering, 197, 107981.

[27] Lungu, C. (2020). Machine Learning Blog: Training an RNN with teacher forcing. Retrieved from http://www.clungu.com/Teacher-Forcing/