VOLUME 17, N° 2 2023
Journal of Automation, Mobile Robotics and Intelligent Systems

Cuban Consumer Price Index Forecasting Through Transformer with Attention

Submitted: 10th January 2023; accepted: 1st August 2023

Reynaldo Rosado, Orlando G. Toledano-López, Hector R. González, Aldis J. Abreu, Yanio Hernandez

DOI: 10.14313/JAMRIS/2-2023/11

Abstract:
Recently, time series forecasting modelling in the Consumer Price Index (CPI) has attracted the attention of the scientific community. Several research projects have tackled the problem of CPI prediction for their countries using statistical learning, machine learning and deep neural networks. The most popular approach to CPI in several countries is the Autoregressive Integrated Moving Average (ARIMA) due to the nature of the data. This paper addresses the Cuban CPI forecasting problem using Transformer with attention model over univariate dataset. The fine tuning of the lag parameter shows that Cuban CPI has better performance with small lag and that the best result was in $p = 1$. *Finally, the comparative results between ARIMA and our proposal show that the Transformer with attention has a very high performance despite having a small data set.*

Keywords: *Consumer price index, Time series forecasting, Transformer with attention, ARIMA, LSTM.*

1. Introduction

The Consumer Price Index (CPI) is a macroeconomic indicator that aims is to measure the variations over time of the prices of goods and services, included in the family basket, which respond to the final consumption expenditures of households. The CPI, one of the most popular indicators in the field of social and economic statistics, is a general macroeconomic indicator and a reference for measuring inflation [24]. The main elements for the construction of the index are: having a representative basket of household expenditures and a weighting structure that defines the importance of each of these products in the population's consumption. The weighting assigned to each good or service determines the effect that the variation in its price will have on the CPI [20].

In the case of Cuba, the weighting reflects the data obtained in the National Survey of Household Income and Expenditures (ENIGH), which was conducted between August 2009 and February 2010. The weights of goods and services are therefore based on the consumption expenditures that households have access to at that time. The goods and services that affect Cuba's CPI are: 01 Food and non-alcoholic beverages; 02 Alcoholic beverages and tobacco; 03 Clothing; 04 Housing services; 05 Furniture and household items; 06 Health; 07 Transportation; 08 Communications; 09 Recreation and culture; 10 Education; 11 Restaurants and hotels; 12 Miscellaneous personal care goods and services [5].

The Monthly Publication of the CPI from the National Office of Statistics and Information (ONEI), gives the average variation experienced by the prices of a basket of goods and services, representative of the consumption of the population in a given period. Approximately 33596 prices are collected monthly, in 8607 establishments, located in 18 municipalities throughout Cuba, the urban area of the head municipalities of 14 provinces and 4 municipalities of Havana province, obtaining national coverage. This means that the index to be shown is only representative of the country; it does not exist at the level of regions or municipalities. The basket of goods and services includes 298 items that represent more than 90.0% of household expenditure. The data are published in the form of reports in pdf format, which makes it difficult to process and analyze them because there is no integrated view of the database [18].

Banking, financial and government authorities systematically monitor the behavior of this indicator as a measure of inflation. It is a reference for monetary policy decisions in all countries. Public administrations frequently analyze this indicator to evaluate issues such as retirement, unemployment, average wage, subsidies [9]. Due to the nature of the data and the frequency with which it is captured, the historical collection of data is considered to be univariate time series as a global metric CPI or multivariate where the goods and services are taken into account.

In general, several approaches in the CPI forecasting field, model the problem as a univariate time series, concentrating only on the study of the global indicator. Approaching it as a multivariate problem, taking into account the variation of the prices of each goods or service included in the basket, does not work well since the global index is composed of the weighted aggregation of the prices of each products. The most widely used statistical method for forecasting the CPI as a univariate time series problem has been the family of the Autoregressive Integrated Moving Average (ARIMA) [1, 6, 7, 15–17]. Recently, deep learning techniques for time series forecasting have improved the performance of CPI prediction. Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM) architectures have the ability to capture time dependence in data, while handling more than one output variable to estimate more than one time instant. Three examples that show good performance with simple LSTM [26] model and temporal data at different time intervals are from Mexico [8], Ecuador [19, 21] and Indonesia [13].

A considerable number of studies were done to analyse the CPI data in different countries. For CPI forecasting, ARIMA models have been widely employed due to the nature of this type of time series, which have very few data samples. The evaluation methodology contemplates the analysis of the seasonality of the time series and the evaluation and selection of the best forecasting model, following the exhaustive strategy of the walk forward validation. On the other hand, some works have focused on the use of deep neural network models. In this type of approach, the great challenge is to obtain generalized models with high accuracy on small data samples, as in the case of the CPI time series.

The aim of this work is to develop a new dataset of Cuba's monthly CPI and a respective time series forecasting model based in Attention mechanism that considers a mapping to a large data representation in encode-decode architecture. In the deep learning proposal, a set of attributes must be represented with an adequate treatment of the non-linearity in the time series relationship. On the other hand, the model should consider a model selection over grid search and a fine tuning of the parameters in the learning process. A comparative baseline study was developed taking in to account ARIMA, and our proposal based on Transformer with attention.

2. Basic Concepts and Notation in Time Series Forecasting

A time series is denoted as a collection of values of a given variable or set of variables ordered chronologically and sampled at constant time intervals. It is mathematically defined as a set of random variables $y_{i,t}, t = 1, 2, \dots, n$ that describe a physical phenomenon. Time series forecasting models predict future values of a target $\{y_{i,t+l}\}, l = 1, 2, \dots, m$ for a given variable i at the time t. In the simplest case the one-step-ahead forecasting model and univariate data, we can define the model as:

$$\hat{y}_{t+1} = f(y^j_t, y^j_{t-1}, y^j_{t-2}, \dots, y^j_{t-k}, W^j_k) \qquad (1)$$

Where $\hat{y}_{t+1}$ is the forecast for univariate data at instant $t + 1$, $\{y^j_t, y^j_{t-1}, y^j_{t-2}, \dots, y^j_{t-k}\}$ are the feature representation space of data that consider a lag window of size k, finally W^j_k are the parameters in the model. For those problems in which there is more than one variable, spatially related and whose nature individually shows a temporal relationship, we say that the problem is a multivariate time series. Classical statistical or machine learning models need to consider the univariate or multivariate problem differently, however, deep learning models can handle both with high accuracy [10].

Time series are usually characterized by three components: trend, seasonality and residuals. In real world time series, and particularly in the CPI problem, seasonality can be affected by external agents such as the economic and financial crisis, the prices of the main products in the world market and emerging situations such as the COVID-19 pandemic.

There are many methods that can be used for time series forecasting and there is no seemingly best solution for any type of problem. The choice of model should always depend on the data and the nature of the problem to be solved. Some models may be more robust against outliers, but have worse performance; the best choice depend on the use case or nature of the data. Given the diversity of time series problems across various domains, numerous neural network architectures choices have emerged that obtaining very good results. While in classical statistical models such as autoregressive models (ARIMA), feature engineering is performed manually and often some parameters are optimized also considering the domain knowledge, Deep Learning models learn features and dynamics directly from the data in an autonomous way, learning more complex patterns and capturing the non-linearity in the data [2].

3. Deep Learning for Time Series Forecasting

The most popular architecture in Deep Learning for Time Series Forecasting are the Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Encoder-Decoder Model with attention. In a recent review article [11], LSTM model for time series forecasting is discussed in more detail. The main contribution of this model to recurrent architectures such as RNNs [14, 25] is in the solution of the optimization problem, where classical activation functions tend to gradient vanishing in interactive propagation to capture long-term dependence. The Gated Recurrent Unit (GRU) [3] is the newest generation of RNNs and is quite similar to an LSTM. The main difference between a GRU and an LSTM is that a GRU has gate, an update, and reset gate; while an LSTM has three gates: an input, a forget, and an output gate, which allow for changes in the state vector of a cell while capturing the long-term temporal relationship. When the time series is small, GRU is suggested; on the other hand, if the series is large, it must be LSTM. GRU checks in each iteration and can be updated with short-term information, however LSTM limits the change gradient in each iteration and in this way does not allow the past information to be completely discarded. This is why LSTM is mostly used for 9 long-term dependency modeling. In [11] he states that there are no significant advantages with respect to the computation time of GRU over LSTM, although it has a smaller number of parameters in the cells.

In RNN, LSTM and GRU each input corresponds to an output for the same time step. However in many real problems it is necessary to predict an output sequence given an input sequence of different length.

This scenario is called sequence to sequence mapping model [22], and is behind numerous commonly used applications like forecast a time vector [4, 27]. Attention mechanism gives good results also in the presence of long or small input sequences, as limits cases, which are related to the encode mechanism that controls the size of the representation space. Attention mechanism has also the advantage of being more interpretable that other Deep Learning models, that are generally considered as black boxes since they do not have the ability to explain their outputs. Particularity, in case of CPI where the the data are composed of small sequences, the encode-decoder attention model could be a very good contribution and uncharted area in the scientific literature.

4. Forecasting via Transformer

For forecasting CPI, a Transformer-based architecture with Self-Attention mechanisms is applied. The proposal follows the basic fundamental design proposed in [23], with encoder-decoder. However, the decoder is conceived in a hidden layer having as input h_{last}, which represents the output of the last hidden state of the encoder block, and as output a single neuron.

Figure 1 represents the design of the proposed architecture based on Transformer. At the input to the model, for the feature representation space y we apply a positional encoder layer to process the input to the encoder block. The residual connections of the above mentioned layer are used in the normalization layer. In the encoder block we define the number of heads in the Multi-Head Attention models as $n_{heads} = 10$.

To obtain the results, $f_0 : \mathbb{R}^{mt} \rightarrow \mathbb{R}^{d}$ is considered as an unknown function, where m is the dimension of random variables. The forecasting capability of an estimator $\hat{f}$ of f_0 is measured via $\mathbb{E}D(\hat{f})$:

$$D(\hat{f}) = \frac{1}{m}|y_{t+1} - \hat{f}(y_t^j, y_{t-1}^j, y_{t-2}^j, \ldots, y_{t-k}^j, W_k^j)|_2^2 \quad (2)$$

We fit the internal parameters of the model using AdamW [12] as optimizer with 50 epochs. We perform the hyper-parameter tuning via Grid Search cross-validation, considering the variables step-size

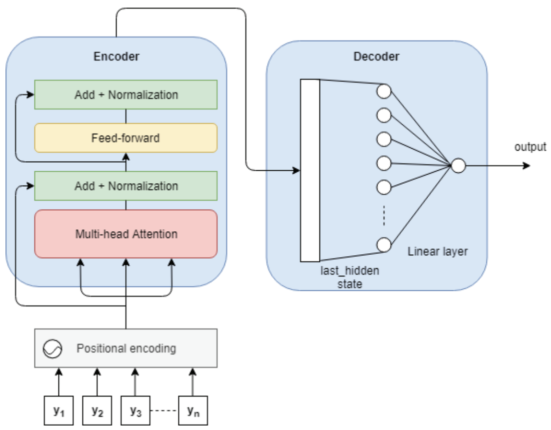

Figure 1. Transformer-based architecture design

Figure 2. Cuba CPI 2010-2021

LR and lag size p. Nevertheless, we manually set the hyper-parameter dimension of the embedding $d_{model} = 100$ on the encoder layer.

5. Results and Discussion

5.1. Dataset

The Cuban Consumer Price Index database was collected from the official website National Office of the Statistic and Information ONEI [18]. This is a monthly time series from January 2010 to December 2020 with very low variability in the data as we can show in Table 1.

The values are the the global Cuban CPI averages for 11 grouped of the categories and almost 298 goods and services. It is necessary to clarify that the data sets in the context of the CPI are very short series; learning models that require a lot of data are not effective in this context. Under these conditions we are modelling an appropriated problem as a time series forecast. Figure 2 shows the trends of the series and seasonality.

The Augmented Dickey-Fuller (ADF) test is a powerful tool used to check the stationarity of the time series. This test can help to choose various parameters such as the size of the windows or the differential order to transform into stationary. The null hypothesis of the ADF test is that the time series is non-stationary. Therefore, if the p-value of the test is below the significance level (0.05), the null hypothesis is rejected and it follows that the time series is truly stationary. In our time series the result of the ADF test can be found in Table 2. The test result shows that the series is non-stationary while the first differential its stationary.

Table 1. Characteristic of the Cuban Consumer Price Index dataset

	CPI
mean	103.35
std	1.37
min	100.12
25%	102.66
50%	103.15
75%	103.89
max	109.5

Table 2. ADF Test over series and the first differential. Lag order d=1

ADF Test, Cuban CPI	**ADF Statistic**	0.154
	p-value	9.7e-01
	Critical Values	
	1%:	-3.488
	5%:	-2.887
	10%:	-2.580
ADF Test, First differential of the Cuban CPI	**ADF Statistic**	-4.089
	p-value	1.01e-03
	Critical Values	
	1%:	-3.488
	5%:	-2.887
	10%:	-2.580

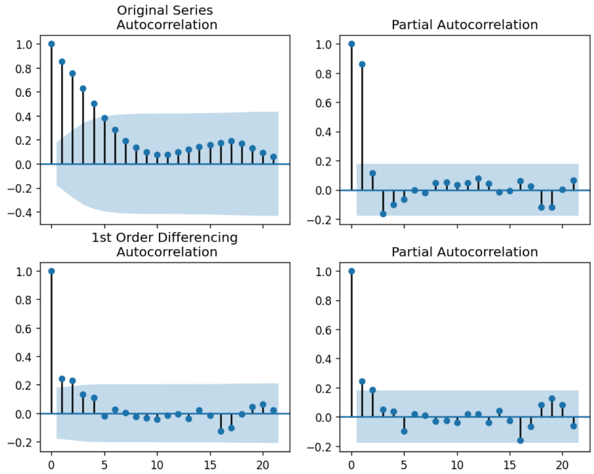

Figure 3. Autocorrelation and Partial Autocorrelation in Cuban CPI

Autocorrelation refers to how correlated a time series is with its past values whereas the Autocorrelation Function (ACF) is the plot used to see the correlation between the points, up to and including the lag unit. Furthermore, the Partial Autocorrelation (PACF) at lag K is the correlation that results after removing the effects of any correlations due to the terms at shorter lags. Figure 3; shows the ACF and PACF of the Cuban CPI for the original and first differential of the series. As we can see the significance in the ACF its defined for a lag $p \leq 4$ while for the PACF is $q \leq 2$.

5.2. Forecasting Measures

As in other similar papers, we use the most common metrics for CPI time series forecasting. The Root Mean Squared Error (RMSE), Mean absolute Error (MAE) and Mean absolute Percentage Error (MAPE). Given a test set with a window size to N, D_{test} observations and a forecast vector $\hat{y}_{t+l} = f(y_{t+l})$, these measures are given as:

$$RMSE(f; D_{test}) = \sqrt{\frac{\sum_{y_{t+l} \in D_{test}} \left(f(y_{t+l}) - \vec{y}\right)^2}{N}} \quad (3)$$

$$MAE(f; D_{test}) = \frac{1}{N} \sum_{y_{t+l} \in D_{test}} \left|\left(f(y_{t+l}) - \vec{y}\right)\right| \quad (4)$$

$$MAPE(f; D_{test}) = \frac{1}{N} \sum_{y_{t+l} \in D_{test}} \left|\frac{\left(f(y_{t+l}) - \vec{y}\right)}{\vec{y}}\right| \quad (5)$$

5.3. Validation and Results

In order to validate the model the transformer we define the follow methodology in the validation. Firstly, the data were split into training and testing, using data prior to January 1, 2018 to train the models and the last two years to validate the model. The first experiment was aimed to select the best model using a grid search of the parameters in combination with the fine tuning. The parameters, fine tuning is evaluated at two levels, one of them related to the time series variable representation in the input layer and in the second level we consider the internal parameters of the transformer. In all cases, different runs were performed with combinations of the parameter settings to obtain the best models at each levels. The testing strategy in case of the transformer model, considers a "multi step ahead" approach procedure while in case of the statistic model like ARIMA, considered in the comparison of the challenge, use "one walk-forward" procedure, which learns one model in each step.

Figure 4 shows the behaviour of the different metrics studied with a set of possible lag sizes $p = \{1, 2, 3, 6, 9, 12\}$. The best solution can be found in $p = 1$ for all metrics. This result is in correspondence with ADF test and statistic analysis developed in the previous section. A second level of parameter tuning was developed internally towards the choice of step size in the AdamW adaptive gradient optimization method. For this, we chose a set of parameters in the range $LR = \{0.00027, 0.00025, 0.00023, 0.00021, 0.00019\}$. The plot of the behavior of the error metrics shows little sensitivity of the models to the parameter setting, which corresponds to the robustness of the adaptive

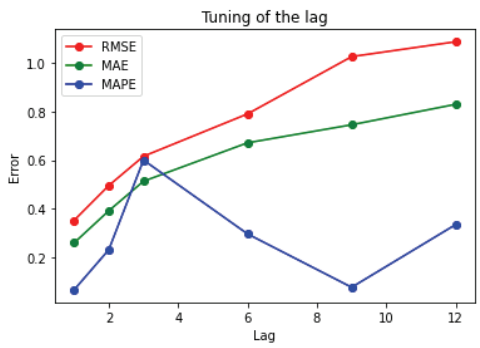

Figure 4. Fine tuning of the lag size parameter

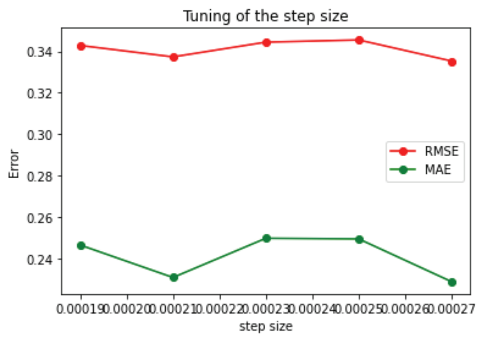

Figure 5. Fine tuning of the step size parameter

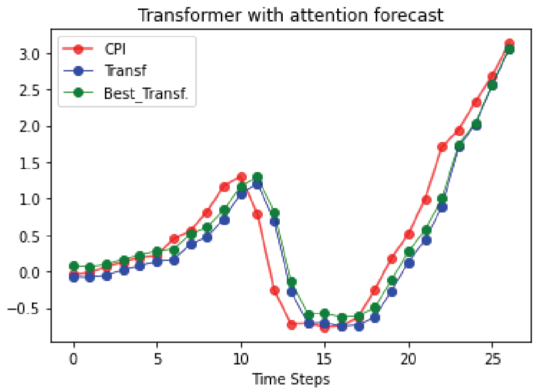

Figure 6. Forecasting based Transformer

Table 3. Challenge of the proposal and ARIMA models

	Parameters	RMSE	MAE	MAPE
ARIMA (1,2,1)	$(p, d, q) = (1, 2, 1)$	0.714	0.433	**0.004**
ARIMA_BoxJenk	$(p, d, q) = (1, 2, 1)$	0.657	0.426	**0.004**
Persistence	-	0.739	0.530	0.005
Transformer	AdamW, $LR = 0.00025$	0.346	0.249	0.717
Best_Transformer	AdamW, $LR = 0.00021$	**0.337**	**0.231**	0.631

gradient as a solver method. Los resultados de la prediccion de los ultimos dos anos a partir del mejor modelo se ilustra en la figura 6, con un buen ajuste del modelo. Similarly, in the experimentation, several ARIMA performances were carried out in order to compare the results of modeling the problem via transformer with respect to the classical approaches used in similar works. For the ARIMA evaluation, different combinations of the parameters (p, d, q) were executed, obtaining $(1, 2, 1)$ as the best combination. In addition, a nonlinear processing based on a Box Jenkins function representation space was included and finally a persistence forecasting model using the previous months was used, which allows us to define the worst-case forecasting scenario. Table 3 summarizes the results of the models studied in an integrated view. The proposed solution for CPI forecasting in Cuba shows superior results for the all metrics in respect to the classical ARIMA models. This result makes us reflect on the potentialities of deep learning and in particular the transformer-based models with attention mechanisms which have demonstrated in this case study their superiority over the classical approaches. In future studies it is suggested to extend the case studies to the CPI of other countries and to consider the multivariate problem.

6. Conclusion

From the tests performed, it can be concluded that the univariate models for the study of the Cuban CPI behave with high efficiency under the conditions of the evaluation considered and the introduction of the Transformer with attention. The good results of the deep neural networks and particularly the Transformers show stables results in the CPI dataset conditions (Univariate and very short time series). This result its not common in deep learning in general, where the models need very large samples of data. On the other hand, the results evidenced in this work establish a baseline for the CPI database in Cuba and at the same time provide a reference to be used in data sets from other countries.

AUTHORS
Reynaldo Rosado – Universidad de las Ciencias Informáticas, Cuba, e-mail: rrosado@uci.cu.
Orlando G. Toledano-López – Universidad de las Ciencias Informáticas, Cuba, e-mail: ogtoledano@uci.cu.
Hector R. González – Universidad de las Ciencias Informáticas, Cuba, e-mail: hglez@uci.cu.
Aldis J. Abreu – Universidad de las Ciencias Informáticas, Cuba, e-mail: ajabreu@uci.cu.
Yanio Hernandez – Universidad de las Ciencias Informáticas, Cuba, e-mail: yhernandezh@uci.cu.

ACKNOWLEDGEMENTS

This work was supported by project PN223LH006013 *Plataforma para el análisis de grandes volúmenes de datos y su aplicación a sectores estratégicos*.

References

[1] A. Banerjee. "Forecasting price levels in india–an arima framework", *Academy of Marketing Studies Journal*, vol. 25, no. 1, 2021, 1–15.

[2] J. Brownlee, *Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python*, Machine Learning Mastery, 2018.

[3] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling". In: *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.

[4] S. Du, T. Li, Y. Yang, and S.-J. Horng. "Multivariate time series forecasting via attention-based encoder–decoder framework", *Neurocomputing*, vol. 388, 2020, 269–279.

[5] J. M. García Molina, *La economía cubana a inicios del siglo XXI: desafíos y oportunidades de la globalización*, CEPAL, 2005.

[6] A. Ghazo et al. "Applying the arima model to the process of forecasting gdp and cpi in the jordanian economy", *International Journal of Financial Research*, vol. 12, no. 3, 2021, 70.

[7] S. Jere, A. Banda, R. Chilyabanyama, E. Moyo, et al. "Modeling consumer price index in zambia: a comparative study between multicointegration and arima approach", *Open Journal of Statistics*, vol. 9, no. 02, 2019, 245.

[8] L. M. León Anaya, V. M. Landassuri Moreno, H. R. Orozco Aguirre, and M. Quintana López. "Predicción del ipc mexicano combinando modelos econométricos e inteligencia artificial", *Revista mexicana de economía y finanzas*, vol. 13, no. 4, 2018, 603–629.

[9] J. Li, Y. Vidyattama, H. A. La, R. Miranti, and D. M. Sologon. “Estimating the impact of covid-19 and policy responses on australian income distribution using incomplete data”, *Social Indicators Research*, vol. 162, no. 1, 2022, 1–31.

[10] B. Lim, and S. Zohren. “Time-series forecasting with deep learning: a survey”, *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, 2021, 20200209.

[11] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich. “A survey on long short-term memory networks for time series prediction”, *Procedia CIRP*, vol. 99, 2021, 650–655.

[12] I. Loshchilov, and F. Hutter. “Decoupled weight decay regularization”, *arXiv preprint arXiv:1711.05101*, 2017.

[13] D. P. Manik et al. “A strategy to create daily consumer price index by using big data in statistics indonesia”. In: *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2015, 1–5.

[14] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur. “Recurrent neural network based language model”. In: *Eleventh annual conference of the international speech communication association*, 2010.

[15] J. Mohamed. “Time series modeling and forecasting of somaliland consumer price index: a comparison of arima and regression with arima errors”, *American Journal of Theoretical and Applied Statistics*, vol. 9, no. 4, 2020, 143–53.

[16] T. Nyoni. “Modeling and forecasting inflation in kenya: Recent insights from arima and garch analysis”, *Dimorian Review*, vol. 5, no. 6, 2018, 16–40.

[17] T. Nyoni. “Arima modeling and forecasting of consumer price index (cpi) in germany”, 2019.

[18] ONEI. “Índice de precios al consumidor base diciembre 2010”, 2022.

[19] J. Riofrío, O. Chang, E. Revelo-Fuelagán, and D. H. Peluffo-Ordóñez. “Forecasting the consumer price index (cpi) of ecuador: A comparative study of predictive models”, *International Journal on Advanced Science, Engineering and Information Technology*, vol. 10, no. 3, 2020, 1078–1084.

[20] R. A. G. Rodríguez. “La visión estructuralista de la inflación. primeras aproximaciones para cuba.”, 2017.

[21] R. Rosado, A. J. Abreu, J. C. Arencibia, H. Gonzalez, and Y. Hernandez. “Consumer price index forecasting based on univariate time series and a deep neural network”. In: *International Workshop on Artificial Intelligence and Pattern Recognition*, 2021, 33–42.

[22] I. Sutskever, O. Vinyals, and Q. V. Le. “Sequence to sequence learning with neural networks”, *Advances in neural information processing systems*, vol. 27, 2014.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. “Attention is all you need”, *Advances in neural information processing systems*, vol. 30, 2017.

[24] W. Wibowo, T. Purwa, E. N. A. Bahri, B. S. S. Ulama, and R. N. Wilantari. “Impacts of earthquakes on consumer price index and inflation: A case study in west nusa tenggara province, indonesia”. In: *Journal of Physics: Conference Series*, vol. 1863, no. 1, 2021, 012062.

[25] R. J. Williams, and D. Zipser. “A learning algorithm for continually running fully recurrent neural networks”, *Neural computation*, vol. 1, no. 2, 1989, 270–280.

[26] S. Zahara, M. Ilmiddaviq, et al. “Consumer price index prediction using long short term memory (lstm) based cloud computing”. In: *Journal of Physics: Conference Series*, vol. 1456, no. 1, 2020, 012022.

[27] B. Zhang, G. Zou, D. Qin, Y. Lu, Y. Jin, and H. Wang. “A novel encoder-decoder model based on read-first lstm for air pollutant prediction”, *Science of The Total Environment*, vol. 765, 2021, 144507.