UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Matemáticas y Computacionales

TÍTULO: Forecasting the Consumer Price Index (CPI) of Ecuador: A Comparative Study of Predictive Models

Trabajo de integración curricular presentado como requisito para la obtención del título de Ingeniero en Tecnologías de la Información

Autor:

Riofrío Valarezo Juan Fernando

Tutor:

Ph.D Chang Tortolero Oscar Guillermo

Urcuquí, Octubre 2020

Urcuquí, 28 de septiembre de 2020

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS MATEMÁTICAS Y COMPUTACIONALES
CARRERA DE TECNOLOGÍAS DE LA INFORMACIÓN
ACTA DE DEFENSA No. UITEY-ITE-2020-00031-AD

A los 28 días del mes de septiembre de 2020, a las 15:00 horas, de manera virtual mediante videoconferencia, y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. AMARO MARTIN, ISIDRO RAFAEL , Ph.D.
Miembro No Tutor	Dr. INFANTE QUIRPA, SABA RAFAEL , Ph.D.
Tutor	Dr. CHANG TORTOLERO, OSCAR GUILLERMO , Ph.D.

El(la) señor(ita) estudiante RIOFRIO VALAREZO, JUAN FERNANDO, con cédula de identidad No. 1722042486, de la ESCUELA DE CIENCIAS MATEMÁTICAS Y COMPUTACIONALES, de la Carrera de TECNOLOGÍAS DE LA INFORMACIÓN, aprobada por el Consejo de Educación Superior (CES), mediante Resolución RPC-SO-43-No.496-2014, realiza a través de videoconferencia, la sustentación de su trabajo de titulación denominado: FORECASTING CONSUMER PRICE INDEX (CPI) OF ECUADOR: A COMPARATIVE STUDY OF PREDICTIVE MODELS, previa a la obtención del título de INGENIERO/A EN TECNOLOGÍAS DE LA INFORMACIÓN.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. CHANG TORTOLERO, OSCAR GUILLERMO , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación a través de videoconferencia, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Presidente Tribunal De Defensa	Dr. AMARO MARTIN, ISIDRO RAFAEL , Ph.D.	10,0
Miembro Tribunal De Defensa	Dr. INFANTE QUIRPA, SABA RAFAEL , Ph.D.	10,0
Tutor	Dr. CHANG TORTOLERO, OSCAR GUILLERMO , Ph.D.	10,0

Lo que da un promedio de: 10 (Diez punto Cero), sobre 10 (diez), equivalente a: APROBADO

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

RIOFRIO VALAREZO, JUAN FERNANDO
Estudiante

Dr. AMARO MARTIN, ISIDRO RAFAEL , Ph.D.
Presidente Tribunal de Defensa

Dr. CHANG TORTOLERO, OSCAR GUILLERMO , Ph.D.
Tutor

AUTORÍA

Yo, **JUAN FERNANDO RIOFRÍO VALAREZO**, con cédula de identidad 1722042486, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autora (a) del trabajo de integración curricular, ensayo o artículo científico. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Octubre 2020.

Juan Fernando Riofrío Valarezo
CI: 1722042486

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **JUAN FERNANDO RIOFRÍO VALAREZO,** con cédula de identidad 1722042486, cedo a la Universidad de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular, en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior. En caso que el ensayo o artículo se encuentre aprobado para publicación en bases indexadas, únicamente se publicará el resumen del mismo.

Urcuquí, Octubre 2020.

_____
Juan Fernando Riofrío Valarezo
CI: 1722042486

Dedicatoria

A mi amada madre Greta Valarezo que siempre me impulsó a seguir adelante en mis estudios, a mi amado padre Felipe Riofrío que siempre me apoyó en cualquier decisión tomada. Les debo todo a ellos 2, esta es una pequeña muestra del amor que les tengo y de lo importante que son en mi formación como persona e ingeniero.

Juan Fernando Riofrío Valarezo

Agradecimiento

Para Oscar Chang y Diego Peluffo, grandes personas y profesores. Sin su ayuda no hubiera logrado tan importante logro, ellos me ayudaron a iniciarme en el camino de la ciencia e investigación ayudándome a publicar mi primer artículo científico.

A mis profesores de la Universidad Yachay Tech, cada uno aportó importantes conocimientos y varios me ayudaron a ser una mejor persona. Un especial agradecimiento a Juan Lobos, Juan Mayorga, Rigoberto Fonseca, Fredy Cuenca, Israel Pineda y Zenaida Castillo que me enseñaron mucho más que una simple materia, me ayudaron a ser mejor en muchos sentidos con sus enseñanzas.

A mi familia, especialmente a mis abuelos: Inés, Luis, Miguel y Nancy que siempre creyeron en mí. Gracias a ellos, a mis padres y mis hermanas soy la persona que soy ahora. Gracias a sus valores y conocimientos pasados a mí persona he podido avanzar en mi vida, intentando ser siempre mejor.

Finalmente, a mis amigos tanto de la universidad como del colegio, quienes fueron una parte fundamental de los logros alcanzados durante esta etapa de mi vida, especialmente a Ricardo Vera. Todos siempre estuvieron presentes y siempre conté con su apoyo.

Juan Fernando Riofrío Valarezo

Resumen

El Indice de Precios al Consumidor (IPC) es uno de los indicadores económicos más importantes para la caracterización de los países y generalmente se considera una medida oficial de inflación. El IPC tiene en cuenta la variación mensual de precios de un conjunto determinado de bienes y servicios en una región específica y es clave en la planificación económica y social de un país determinado, de ahí la gran importancia de la predicción del IPC. En este trabajo, se describe un estudio comparativo de modelos predictivos de última generación sobre un conjunto de datos del IPC ecuatoriano con 174 registros mensuales, de 2005 a 2019. Este pequeño conjunto de datos disponible hace que el pronóstico sea una tarea de predicción de series de tiempo desafiante. Otra dificultad es la variación de tendencia de los años pasados, que desde mediados de 2016 ha cambiado de un promedio claramente ascendente de 3.5 puntos a una tendencia estable de ±0.8 puntos. Este documento explora el desempeño de los modelos predictivos relevantes al abordar el problema de pronóstico del IPC ecuatoriano con precisión durante los próximos 12 meses. Para esto, se lleva a cabo un estudio comparativo que considera una variedad de modelos predictivos, que incluyen: Enfoque de redes neuronales usando el modelo secuencial Long Short-Term Memory, aprendizaje automático usando la regresión de vectores de soporte, así como enfoques clásicos como SARIMA y suavizado exponencial. También consideramos herramientas de grandes corporaciones como Facebook Prophet. Como resultado, el trabajo presenta los mejores modelos predictivos y parametros encontrados, junto con el pronóstico del IPC de Ecuador para los próximos 12 meses (parte de 2020). Esta información podría usarse para la toma de decisiones en varios temas importantes relacionados con las actividades sociales y económicas.

Palabras Clave:

Índice de Precios al Consumidor (IPC); Ecuador; Modelos Predictivos; Pronóstico.

Abstract

The Consumer Price Index (CPI) is one of the most important economic indicators for countries' characterization and is typically considered an official measure of inflation. The CPI takes into account the monthly price variation of a determined set of goods and services in a specific region and it is key in the economic and social planning of a given country, hence the great importance of CPI forecasting. In this paper, we outline a comparative study of state-of-the-art predictive models over an Ecuadorian CPI dataset with 174 monthly registers, from 2005 to 2019. This small available dataset makes forecasting a challenging time-series-prediction task. Another difficulty is last year's trend variation, which since mid-2016 has changed from a clearly upward average of 3.5 points to a stable trend of ±0.8 points. This paper explores the performance of relevant predictive models when tackling the Ecuadorian CPI forecasting problem accurately for the next 12 months. For this, a comparative study considering a variety of predictive models is carried out, including: Neural networks approach using Sequential Model with Long Short-Term Memory layers machine learning using Support Vector Regression, as well as classical approaches like SARIMA and Exponential Smoothing. We also consider big corporations' tools like Facebook Prophet. As a result, the paper presents the best predictive models and parameters found, along with Ecuador's CPI forecasting for the next 12 months (part of 2020). This information could be used for decision-making in several important topics related to social and economic activities.

Forecasting the Consumer Price Index (CPI) of Ecuador: A Comparative Study of Predictive Models

Juan Riofrío[#], Oscar Chang[#], E. J. Revelo-Fuelagán[*], and Diego H. Peluffo-Ordóñez[#]

[#] *Department of Computational Sciences, Yachay Tech University, Urcuquí, 100650, Ecuador.*
E-mail: jriofrio93@gmail.com, dpeluffo@yachayetech.edu.ec, ochang@yachaytech.edu.ec

[*] *Department of Electronic Engineering, Universidad de Nariño, Nariño, 52001, Colombia.*
E-mail: javierrevelof@udenar.edu.co

Abstract— **The Consumer Price Index (CPI) is one of the most important economic indicators for countries' characterization and is typically considered an official measure of inflation. The CPI takes into account the monthly price variation of a determined set of goods and services in a specific region and it is key in the economic and social planning of a given country, hence the great importance of CPI forecasting. In this paper, we outline a comparative study of state-of-the-art predictive models over an Ecuadorian CPI dataset with 174 monthly registers, from 2005 to 2019. This small available dataset makes forecasting a challenging time-series-prediction task. Another difficulty is last year´s trend variation, which since mid-2016 has changed from a clearly upward average of 3.5 points to a stable trend of ±0.8 points. This paper explores the performance of relevant predictive models when tackling the Ecuadorian CPI forecasting problem accurately for the next 12 months. For this, a comparative study considering a variety of predictive models is carried out, including: Neural networks approach using Sequential Model with Long Short-Term Memory layers machine learning using Support Vector Regression, as well as classical approaches like SARIMA and Exponential Smoothing. We also consider big corporations´ tools like Facebook Prophet. As a result, the paper presents the best predictive models and parameters found, along with Ecuador´s CPI forecasting for the next 12 months (part of 2020). This information could be used for decision-making in several important topics related to social and economic activities.**

Keywords— **Consumer Price Index (CPI); Ecuador; Predictive Models; Forecasting**

I. INTRODUCTION

Since Ecuador adopted dollarization in 2000, its economy has shown a strong and persistent development after a big economic crisis that resulted in an annual inflation rate of 91% that year. The following years served to stabilize the country and its economy, as inflation rates began to decrease due to recession and the non-appearance of devaluation [1]. In the last years Ecuador has had stable and low inflation, closing at 0.27% in 2018 and showing a recovery in the economic life of the country [2]. In Ecuador, as in most countries, inflation is measured through the changes in the Consumer Price Index (CPI) between two periods.

The CPI is a monthly record that allows measuring throughout the time the variation of final consumption prices of a determined set of goods and services by households living in a specific urban area. In this sense, it results advisable to systematically change the CPI base year and update the goods and services of the market basket and its weights in order to let the index adjust to market evolution and changes in the population consumption. In Ecuador, six CPI base year changes have occurred. This work was developed with the last change, in 2014 [3]. In this update (Base: 2014 = 100), the CPI goods and services basket consisted of 359 products divided into 12 groups and then into 43 subgroups. Roughly 25,350 product prices increase monthly in 9 cities in Ecuador. This change of the base year was based on the results of the National Survey of Income and Expenditure of Urban and Rural Households (ENIGHUR by its acronym in Spanish) conducted in 2011-2012.

The dataset analysed in this work is the General CPI of Ecuador from 2005 to 2019, which contains 174 samples and the trend. Its small size makes it difficult to learn patterns, since it is prone to over-fitting and holds few data to test

results. The dataset shows a clearly growing trend varying an average of 3.5 points each year until 2016, and a stable trend in the same range (±0.8 points) from then on. The data were taken from the Ecuadorian National Institute of Statistics and Censuses (INEC by its acronym in Spanish).

CPI is a key variable for a country's macroeconomic analysis and it can be helpful when performing several studies. Some clear examples are its use in determining the cost of living in a city, inflation, or the minimum wage of a country. It can also be used to predict gold price returns as proposed by Sharma [4] or to forecast regional socio-economic development [5]. Due to its importance, this study attempts to forecast the monthly values of CPI for the next year (July 2019 – June 2020) using 5 different time series predictive models. First, a machine learning approach was selected, represented by Support Vector Regression (SVR) using different kernels. Second, an Artificial Neural Network (ANN) approach using Long Short-Term Memory (LSTM) Layers was selected. The third predictive model used was the Seasonal Autoregressive Integrated Moving Average (SARIMA), a classical approach used in statistics and econometrics. The next predictive model used was the Exponential Smoothing model. Finally, a predictive method designed by Facebook called Prophet was used as well. All the implementations were done in Python programming language.

The remaining of this paper is structured as follows: Section II describes the dataset used and overviews the predictive models considered as well as the experimental methodology proposed. Section III gathers the results and discussion. Finally, the concluding remarks are drawn in section IV.

II. MATERIALS AND METHODS

A. Dataset

The "Consumer Price Index of Ecuador" dataset is a time series composed of 174 samples, each one has been taken monthly from January 2005 to June 2019. The dataset is shown in Fig. 1.
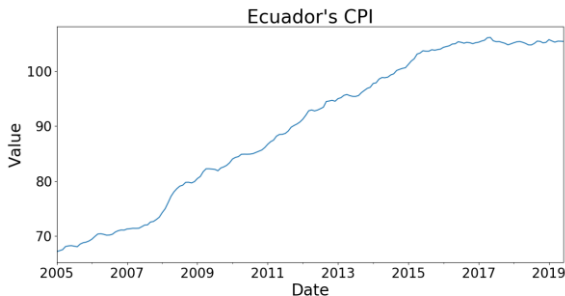

Fig. 1: 2005-20019 Ecuador´s CPI time series plot

The only column of the dataset contains the CPI value for each month. Therefore, it is a vector of dimension 174×1. The dataset, together with the methodology used to calculate the monthly value of CPI, can be found in the official governmental website: https://www.ecuadorencifras.gob.ec/indice-de-precios-al-consumidor/ . Table 1 presents the dataset description.

TABLE I
ECUADOR'S CPI DATASET DESCRIPTION

Count	Mean	Std. Deviation	Min	Max
174	89.957184	13.306077	67.12	106.17

Figures 2, 3 and 4 show the seasonal decomposition of the time series.
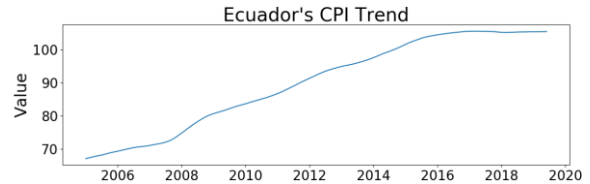

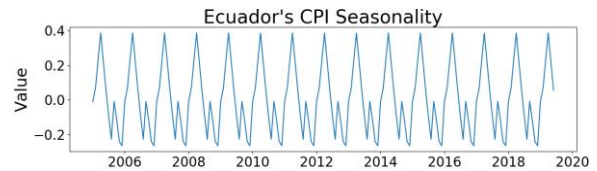Fig. 2: Trend attribute of Ecuador's CPI time series additive seasonal decomposition.

Fig. 3: Seasonality attribute of Ecuador's CPI time series additive seasonal decomposition. Frequency = 12.
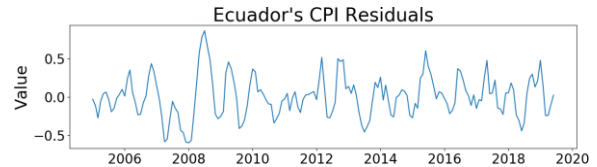

Fig. 4: Residuals attribute of Ecuador's CPI time series additive seasonal decomposition.

B. Predictive Models

The Ecuadorian CPI dataset was used to train and evaluate five different predictive models.

1) Support Vector Regression (SVR) [6]: This regression technique uses the same principle of the Support Vector Machine for classification, except for some differences so that it can be applied to continuous values. The idea is to consider a margin distance epsilon, so that all the samples are in a band around the selected hyperplane that best suits our data; this means, that their distance from the hyperplane is smaller than epsilon. When defining the hyperplane only examples that are more distant from our hyperplane than epsilon are taken. In SVR those samples will be considered as support vectors. SVR uses kernels to map the input into a higher dimensional space where we can find the hyperplane that best suits our data. In this work we use Radial Basis Function (RBF) and Polynomial kernels for the SVR model.

2) Long Short-Term Memory (LSTM) Neural Network [7] [8]: LSTM is a recurrent neural network architecture developed by Sepp Hochreiter and Jürgen Schmidhuberin the mid-'90s in response to the problems presented by

recurrent neural networks. Recurrent neural networks need to monitor states, which is computationally expensive, and they have issues like the vanishing gradient and the exploding gradient. LSTM unit is an approach commonly used to deal with these issues.

LSTM unit is a reflection of computer memory that works in conjunction with a recurrent network. It is utilized to maintain, update, and regulate the states of the network model outside of its typical execution flow. It consists of four main elements: the memory cell and three logistic gates. The memory cell holds information and the logistic gates specify the flow of data inside the LSTM. The input gate writes data into the memory cell; the output gate is in charge of reading data from the information cell and sending it back to the recurrent network, whereas the forget gate decides whether to keep or delete data from the information cell. These gates have similarities with the neurons from a neural network, as they are multiplicative analog sigmoid-activated nodes. By manipulating these gates, a recurrent network can remember what it needs and forget what is no longer useful. The idea is to manipulate values through the gates in order to solve the vanishing gradient and exploding gradient problems. The gates let the network forget the states that are not required anymore, thus reducing the model's computational load radically. LSTM let recurrent networks keep learning over 1000 time steps by maintaining a constant error.

3) Seasonal Auto-Regressive Integrated Moving Average (SARIMA) [9] : Auto-Regressive Integrated Moving Average (ARIMA) is a statistical model that finds patterns from existing data to predict future values. Therefore, future pre- dictions are described by past data and not by independent variables. It is one of the most used methods to forecast univariate time series data. It has 3 hyperparameters to specify (p,d,q), where 'p' is the auto-regression term seen as the lags of previous value, 'd' is the integral (differencing) term for making time series stationary, and the moving average term 'q' is a past error (multiplied by a coefficient).

SARIMA is an extension of the ARIMA model, with the difference that SARIMA supports time series with a seasonal component. In order to make this possible, it adds 4 new hyperparameters (sp,sd,sq, s), where three of them are similar to ARIMA hyperparameters (p,d,q) but they implicate back-shifts of the seasonal period. The last hyperparameter added ('s') is the period of the seasonality, where 12 is the monthly data, 4 is the quarterly data, 0 is no seasonal data, etc.

4) Exponential Smoothing [10]: It is a predictive model for univariate data that can support data with a seasonal component or systematic trend by adding some variations. Exponential smoothing uses the linear combination of past values to forecast future values. The difference with moving average is that exponential smoothing sets exponentially decreasing weights as the observation gets older, which means that more recent values are associated with higher weights so they have more influence in the predicted value. Precisely, it weighs past observations with a geometrically declining ratio.

5) Facebook Prophet [11]: It is a software designed by Facebook's Core Data Science team. The so-named Prophet approach is a time series forecasting method that is easy to use and tune. It is based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. Prophet properly handles missing data, changes in trend, and outliers. This predictive model works best with data that have a strong seasonality and some seasons of historical data. Facebook uses this software in its application for its predictions, but because it is easy to use and tune, it is also being used by amateur data scientists. The Python library used for this model is fbprophet, to see the documentation of the procedure go to https://github.com/facebook/prophet .

C. Metrics

The mean absolute percentage error was used to compare among the efficiency of the predictive. In SARIMA model the AIC estimator was also used.

1) Mean absolute Percentage Error (MAPE) [12] : It is the average of the absolute percentage errors of the predictions. This measure is straightforward as it shows how significant prediction errors are compared to the values of the series in terms of percentage. The smaller the MAPE the better the predictions. MAPE can be expressed as follows:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|Y_t - \bar{Y}_t|}{|Y_t|} \qquad (1)$$

where Y_t is the actual value at time t, $\bar{Y}_t$ is the predicted value at the time t and n is the number of observations.

2) Akaike information criterion (AIC) [13]: AIC approximates the relative quantity of information lost by a selected model, thus, if the model loses less information it is a higher quality model. By adding parameters to the model we get better results but we are also trading off against overfitting and actually losing information about the real inherent pattern. AIC represents a trade-off between an added number of parameters and the error that we are considering by adding said parameters. The model that fits better will have the lowest AIC.

D. Experimental Methodology:

For all the experiments (see the next section), the predictive models were trained using all the data except for the last 12 months as seen in Fig. 5. The last 12 months (test dataset) presented in Fig. 5 were used for testing the models. All experiments were run in Python using already existing modules of predictive models. A grid search was performed to look for the best hyperparameters for each kind of predictive model comparing the possible configurations and selecting the one with the lowest MAPE. This is done for

each predictive model and at the end the forecasting of the next 12 months is performed to compare the best configuration of each predictive model against each other. Fig. 6 summarizes graphically the methodology used in this research.
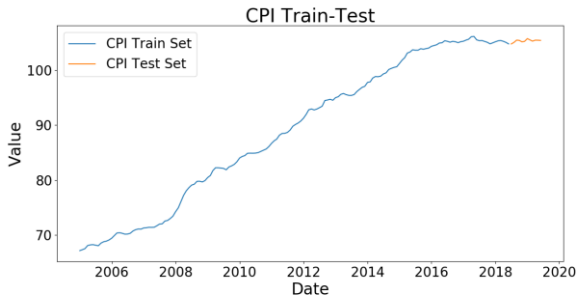

Fig. 5: CPI train & test datasets plot

Fig. 6: Block diagram of the proposed methodology.

III. RESULTS AND DISCUSSION

A. SVR Results

For the SVR model, 2 different kernels were used: RBF and polynomial. For each kernel, a grid search was performed with a range of values to look for the best hyperparameter combination. The implementation was made with the help of the scikit-learn library and the SVR method. The best configurations were the SVR using RBF kernel with an epsilon value of 1, a penalty value of the error of 10, a kernel coefficient of 0.0001; and SVR using polynomial kernel of degree 8 and an independent term value of 9, with an epsilon value of 0.1, a penalty value of the error of 1, and a kernel coefficient of 0.1. After training our models we proceeded to forecast the testing values and find the MAPE for each model. In the case of the RBF kernel, the MAPE was 0.00254, whereas the MAPE for the polynomial kernel was 0.00171. The results of the SVR models are presented in Figures 7, 8, 9, 10 and 11.
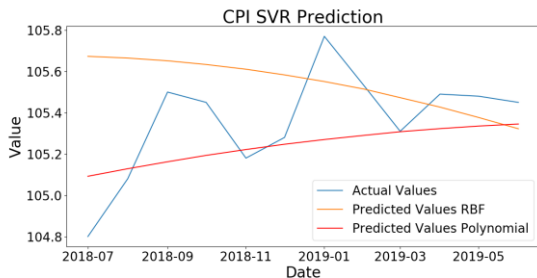
Fig. 7: Comparison of SVR predictions for the test datasets.
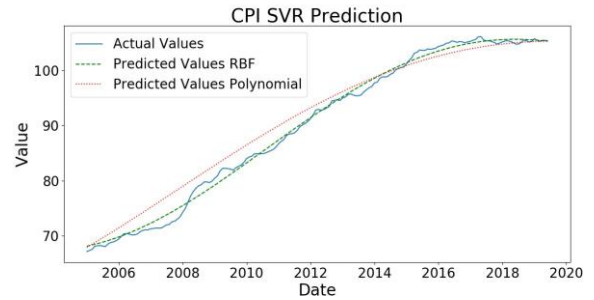

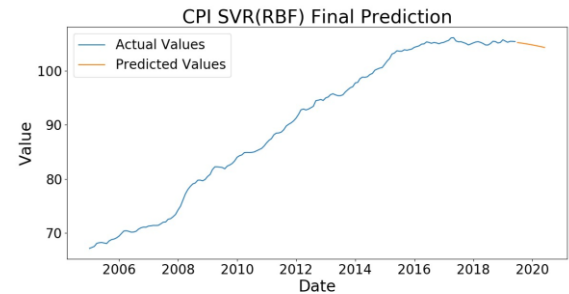Fig. 8: SVR models fitting to the time series.

Fig. 9: CPI dataset and final predictions of SVR model using RBF kernel.
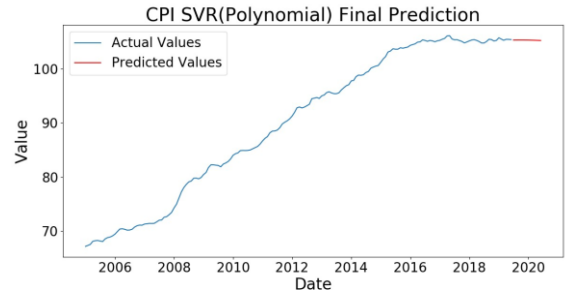

Fig. 10: CPI dataset and final predictions of the SVR model using the polynomial kernel.
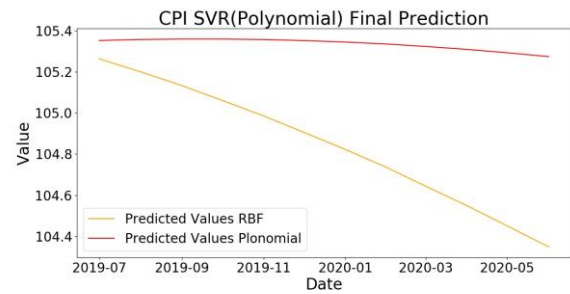

Fig. 11: Comparison of final predictions of SVR models

B. LSTM Neural Network Results

To use this predictive model, we first created an artificial neural network (ANN), which was done with Keras library running on top of TensorFlow. Instead of a grid search for the parameters, we tried different architectures varying the number of layers, as well as the number of units in the layers until obtaining the best possible MAPE. The selected ANN is a sequential model of layers, 3 LSTM layers, 3 Dropout layers and 1 dense layer.

4

The number of units for the LSTM layers was 6 and for the dense layer the unit was 1, the network architecture is presented in Fig. 12.

```
Model: "sequential"

Layer (type)                Output Shape              Param #
=================================================================
lstm (LSTM)                 (None, 12, 6)             192

dropout (Dropout)           (None, 12, 6)             0

lstm_1 (LSTM)               (None, 12, 6)             312

dropout_1 (Dropout)         (None, 12, 6)             0

lstm_2 (LSTM)               (None, 12, 6)             312

dropout_2 (Dropout)         (None, 12, 6)             0

lstm_3 (LSTM)               (None, 6)                 312

dropout_3 (Dropout)         (None, 6)                 0

dense (Dense)               (None, 1)                 7
=================================================================
Total params: 1,135
Trainable params: 1,135
Non-trainable params: 0
```

Fig. 12: LSTM Neural Network architecture summary.

After creating the model, the training and testing stages were done, and in this case, the obtained MAPE was 0.00173. Because the input in this model is different, it is important to mention that in order to forecast a value the selected model uses the past 12 values to predict the next one after training. So to predict the first value we used the past 12 months, added the prediction to the input set and then used it to predict the next values. This method is better to use if you need to predict only one value into the future as it needs the previous values to predict. The results of the LSTM neural network are provided in Figures 13, 14 and 15.
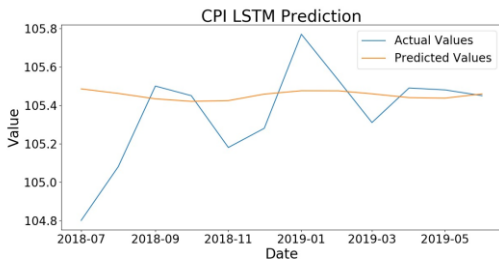

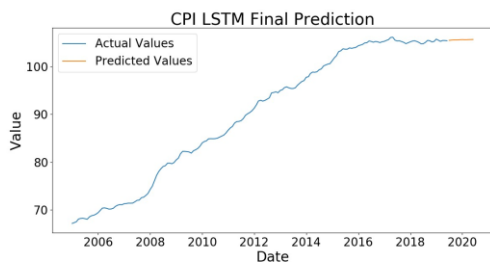Fig. 13: LSTM model predictions for the test datasets

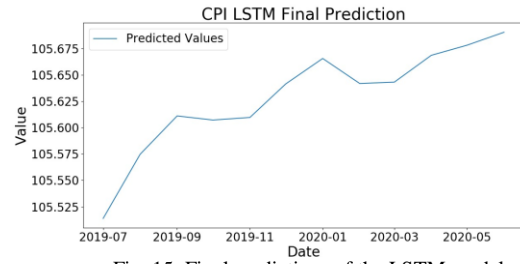Fig. 14: CPI dataset and final predictions of LSTM model.

Fig. 15: Final predictions of the LSTM model.

C. SARIMA Results

In the SARIMA model 2 different metrics were used to test the results of the hyperparameters search. First, the same approach as the other experiments was used, searching for tuning with the less MAPE. In the second approach a search for the hyperparameters to get the best AIC was permormed. We used the method SARIMAX from the library statsmodels to create and train our model. The best configuration to get the lowest MAPE for the SARIMA model was found with the following hyperparameters (p = 2, d = 2, q = 4), (sp = 1, sd = 0, sq = 1, s = 0). A MAPE of 0.00181 was obtained with this configuration. The best hyperparameter to get the best AIC was found with the help of the auto_arima function from the pmdarima library, while the best configuration found for the hyperparameters was (p = 1, d = 1, q = 1), (sp = 0, sd = 1, sq = 1, s = 12). After training and testing the results of the model a higher MAPE of 0.00550 was obtained, as expected. The results of the SARIMA models are presented in Figures 16,17, 18 and 19.
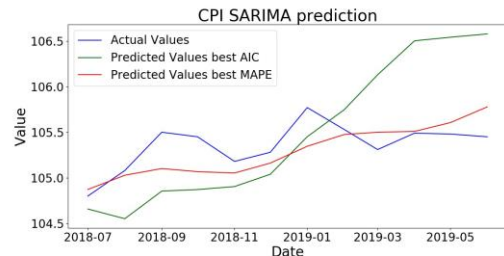

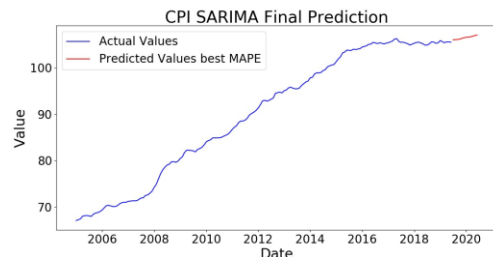Fig. 16: Comparison of SARIMA predictions for the test datasets.

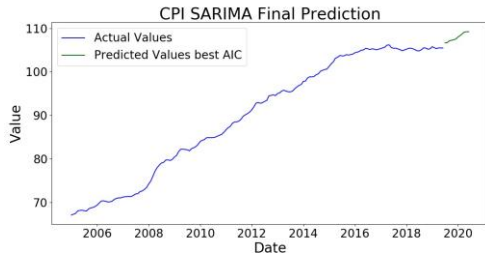Fig. 17: CPI dataset and final predictions of SARIMA (2,2,4)- (1,0,1,0) model.

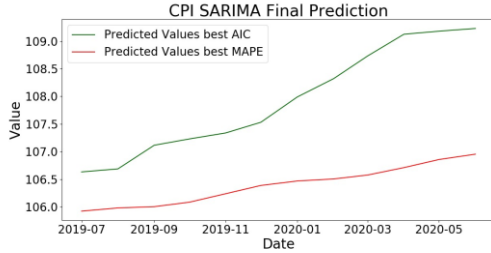Fig. 18: CPI dataset and final predictions of SARIMA (1,1,1)- (0,1,1,12) model.

Fig. 19: Comparison of final predictions of SARIMA models.

D. Exponential Smoothing Results

The model was created using the Exponential Smoothing method from stats model library. The same procedure as before was applied and a grid search was done to look for the best configuration to obtain the lowest MAPE. The best configuration found for the exponential smoothing method was a model using an additive type of trend without damping it, and an additive seasonal component of 12 periods per season.

The predictions obtained with this configuration produced a MAPE of 0.00316 when compared to the test values. All the results of the exponential smoothing model are shown in Figures 20, 21 and 22.
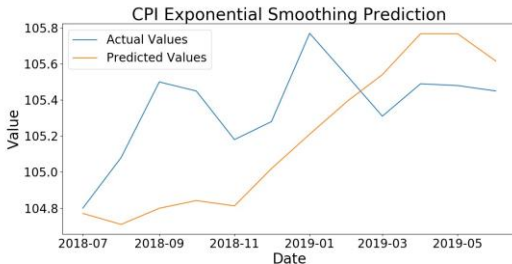

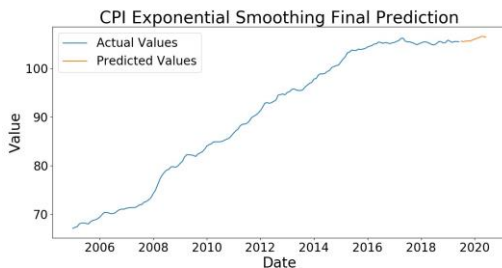Fig. 20: Exponential Smoothing model predictions for test dataset.

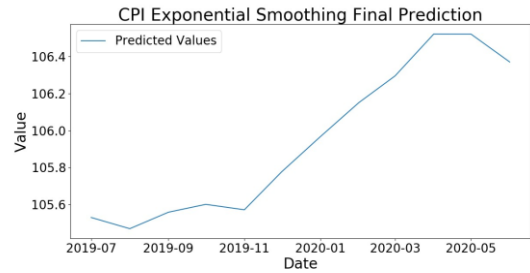Fig. 21: CPI dataset and final predictions of Exponential Smoothing model.

Fig. 22: Final predictions of Exponential Smoothing model.

E. Prophet Results

For the Prophet model, as mentioned before, we used the fbprophet library and the Prophet method. First, we performed the grid search to get the best configuration for the model and used MAPE again to evaluate the best model. The best configuration obtained from this step was the model with linear growth, with an interval width of 0.95 and yearly seasonality.

The model trained with this configuration produces a MAPE of 0.00403, higher than the other methods. The results produced by this model are presented in Figures 23, 24 and 25.
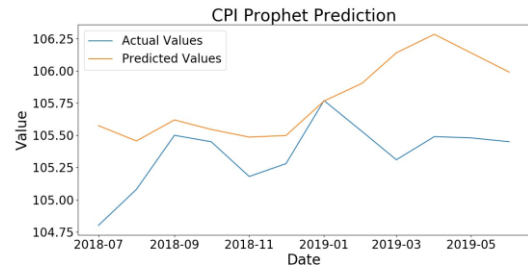

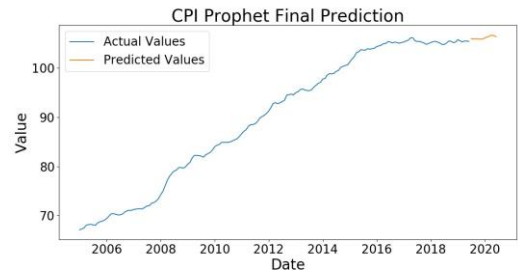Fig. 23: Prophet model predictions for test dataset.

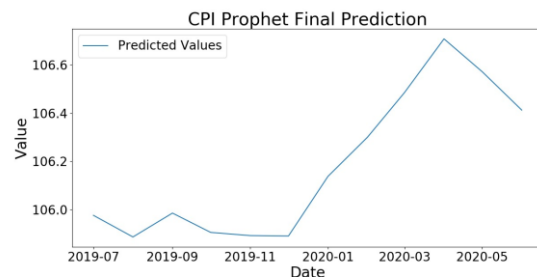Fig. 24: CPI dataset and final predictions of Prophet model.

Fig. 25: Final predictions of Prophet model.

A.   Final Remarks

Finally, Table 2 depicts the final predictions of all the models for the next year.

TABLE III
MONTHLY FINAL PREDICTIONS OF EACH PREDICTIVE MODEL

Month	SVR Poly	SVR RBF	LSTM	SARIMA (MAPE)	SARIMA (AIC)	Exp. Smt	Prophet
2019-07	105.3524	105.2632	105.5138	105.9237	106.6336	105.3726	105.9764
2019-08	105.3572	105.1999	105.5747	105.9816	106.6881	105.4020	105.8868
2019-09	105.3596	105.1325	105.6111	106.0048	107.1160	105.5901	105.9857
2019-10	105.3595	105.0609	105.6072	106.0860	107.2300	105.5512	105.9057
2019-11	105.3570	104.9853	105.6097	106.2392	107.3383	105.4663	105.8925
2019-12	105.3521	104.9057	105.6414	106.3885	107.5314	105.6079	105.8909
2020-01	105.3448	104.8221	105.6656	106.4690	107.9884	105.8691	106.1377
2020-02	105.3352	104.7347	105.6419	106.5060	108.3197	105.8896	106.2999
2020-03	105.3234	104.6436	105.6431	106.5768	108.7298	106.0110	106.4844
2020-04	105.3092	104.5487	105.6686	106.7112	109.1234	106.2835	106.7067
2020-05	105.2928	104.4503	105.6782	106.8578	109.1789	106.2133	106.5705
2020-06	105.2743	104.3483	105.6905	106.9542	109.2281	106.0051	106.4119

Table 3 presents the MAPE obtained in each model when comparing the test dataset with the predictions for the same dates.

TABLE III
PREDICTIVE MODELS' MAPE

Model	MAPE
SVR Poly	0.00171
SVR RBF	0.00254
LSTM	0.00173
SARIMA (MAPE)	0.00181
SARIMA (AIC)	0.00550
Exp. Smoothing	0.00316
Prophet	0.00403

IV. CONCLUSSIONS AND FUTURE WORK

The comparative study of the presented predictive models suggest that the Support Vector Regression (SVR) model using a polynomial kernel is the best element to predict the Ecuadorian CPI one year ahead in the future. The MAPE obtained with this solution was 0.00171. As seen in Figures 7 and 8, it is good in predicting the trend rather than exact values. The model with the second best MAPE was LSTM neural network, which obtained a mark of 0.00173. Interestingly enough the predictions of this model, at least in its graphical representation, look realistic, with peaks and valleys, more like a real CPI should behave. In contrast, predictions with SVR using polynomial (also RBF) tend to follow a smooth curve (Fig. 8). Considering that CPI can be used to estimate inflation, minimum wage or the cost of

living of a region, the results of this work can be used to calculate other important information for the Ecuadorian economy next year.

With the dataset used other alternatives should be taken into account for future works. For example, the modification of a method or combination of two or more methods like the one proposed by Qin et al. to forecast China CPI based on EEMD and SVR method [14]. Another improvement could be to extend the grid to search through more possibilities of hyperparameters, or even better, optimize the search like the one proposed by Cao and Wu [15]for SVR forecasting where the fruit fly optimization algorithm is used to automatically search the best values of the hyperparameters. New values are released every month, so the work of forecasting Ecuador's CPI is a rich source for further researh.

V.   ACKNOWLEDGMENTS

REFERENCES

[1]   M. A. Terán Saltos, "Discursos sobre las causas de la inflación en una economía dolarizada." 2002.
[2]   Banco Central del Ecuador(BCE)., "Ecuador: Reporte mensual de inflación." 2018.
[3]   Instituto Nacional de Estadisticas y Censos(INEC)., "Metodología del Índice de Precios al Consumidor (IPC) Base Anual: 2014 = 100." 2015.
[4]   S. S. Sharma, "Can consumer price index predict gold price returns?," *Econ. Model.*, vol. 55, pp. 269–278, Jun. 2016.
[5]   A. V Babkin, E. P. Karlina, and N. S. Epifanova, "Neural networks as a tool of forecasting of socioeconomic systems strategic development," *Procedia-Social Behav. Sci.*, vol. 207, pp. 274–279, 2015.
[6]   A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3. pp. 199–222, Aug-2004.
[7]   S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[8]   K. Greff, R. K. Srivastava, J. Koutn\'\ik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. neural networks Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, 2016.
[9]   G. E. P. Box and G. M. Jenkins, *Time series analysis forecasting and control*. San Francisco Holden-Day, 1970.
[10]   R. G. Brown and R. F. Meyer, "The Fundamental Theorem of Exponential Smoothing," *Oper. Res.*, vol. 9, no. 5, pp. 673–685, Oct. 1961.
[11]   S. J. Taylor and B. Letham, "Forecasting at scale," *Am. Stat.*, vol. 72, no. 1, pp. 37–45, 2018.
[12]   A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean Absolute Percentage Error for regression models," May 2016.
[13]   S. Hu, "Akaike information criterion," *Cent. Res. Sci. Comput.*, vol. 93, 2007.
[14]   X. Qin, M. Sun, X. Dong, and Y. Zhang, "Forecasting of China Consumer Price Index Based on EEMD and SVR Method," in *Proceedings - 2nd International Conference on Data Science and*

Business Analytics, ICDSBA 2018, 2018, pp. 329–333.

[15]   G. Cao and L. Wu, "Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting," *Energy*, vol. 115, pp. 734–745, Nov. 2016.