

# Prediction model for Peruvian cocoa production using Deep Learning and Long Short-Term Memory techniques

Ronald Porras<sup>1</sup>, Christian Ovalle<sup>2</sup>

<sup>1</sup>Universidad Privada del Norte, Perú, n00201656@upn.pe

<sup>2</sup>Universidad Privada del Norte, Perú, denis.ovalle@upn.pe

*Abstract - Cocoa is an essential product of Peruvian agriculture since a large number of families depend on it. The International Cocoa Organization has suggested carrying out strategies to avoid overproduction and minimize the price crisis. In the research work, the development of a prediction model for the production of Peruvian cocoa was carried out using the techniques of deep learning and lstm, since this will allow estimating production and avoiding the devaluation of the sale price. The PMBOK process guide was used, which has 5 phases, it will also allow control over the risks that may affect the project and greater efficiency in the quality of deliverables. The prediction model for cocoa production was made using a python programming language and the TensorFlow library, which makes it possible to implement models based on deep learning in an easier and faster way. It can be concluded that the prediction model helps to estimate the production of Peruvian cocoa using deep learning and lstm techniques with an accuracy percentage of 99.6%.*

**Keywords.** Prediction model; deep learning; Long short-term memory; Cocoa production, PMBOK

**Digital Object Identifier:** (only for full papers, inserted by LACCEI).  
**ISSN, ISBN:** (to be inserted by LACCEI).  
**DO NOT REMOVE**

# Modelo de Predicción para la producción de cacao peruano utilizando las técnicas de Deep Learning y Long Short-Term Memory

Ronald Porras<sup>1</sup>, Christian Ovalle<sup>2</sup>

<sup>1</sup>Universidad Privada del Norte, Perú, n00201656@upn.pe

<sup>2</sup>Universidad Privada del Norte, Perú, denis.ovalle@upn.pe

## I. INTRODUCCIÓN

**Resumen.** *El cacao es un producto esencial de la agricultura peruana ya que una gran cantidad de familias dependen de ello. La organización internacional del cacao ha sugerido realizar estrategias con el fin de evitar la sobreproducción y minimizar la crisis de precios. En el trabajo de investigación se llevó a cabo el desarrollo de un modelo de predicción para la producción de cacao peruano usando las técnicas deep learning y lstm, ya que esto permitirá estimar la producción y evitar la devaluación del precio de venta. Se utilizó la guía de proceso PMBOK la cual tiene 5 fases, además permitirá tener un control sobre los riesgos que puedan afectar al proyecto y una mayor eficacia en la calidad de los entregables. El modelo de predicción para la producción de cacao se hizo bajo un lenguaje de programación python y la biblioteca TensorFlow, la cual posibilita implementar de una manera más fácil y rápida de modelos basados en deep learning. Se puede concluir que el modelo de predicción ayuda a estimar la producción de cacao peruano utilizando las técnicas de deep learning y lstm con un porcentaje de precisión del 99,6%.*

**Palabras clave.** *Modelo de predicción; Deep learning; Long short term memory; Producción de cacao, PMBOK*

En la actualidad el cacao es un producto importante de la agricultura peruana ya que muchas familias dependen de ello, la mayor cantidad de la producción de cacao proviene de países de bajos ingresos [1]. Según el Ministerio de Desarrollo Agrario y Riego (2020) el cacao tiene un rol esencial en la economía de los sectores agrario e industria. A nivel mundial, el sector de la chocolatería es el principal consumidor de la producción de cacao, también los derivados de sus semillas son usados en la industria de farmacéuticos y cosméticos (manteca de cacao), en dar sabor en bebidas, helados y galletas (cacao en polvo). El origen del cacao peruano está centrado en la amazonía del país, la zona oriente, ya que cuenta con una diversidad genética para obtener así cacao con aromas finos que son apreciados por chocolateros exigentes a nivel mundial. De esta forma, se tiene como las cinco regiones de mayor producción a San Martín, Junín, Ucayali, Huánuco y Cusco durante el año 2019, ocupando así el 84% del total de cacao nacional producido.

Así mismo Romero [2] nos dice que el cacao se cultiva comercialmente en regiones tropicales. "Entre 15 ° al norte de la línea ecuatorial y 15 ° al sur. La temperatura media anual es de 23 ° - 30 ° C, y la temperatura óptima es de 25 ° C. Crece desde el nivel del mar hasta 1200 metros sobre el nivel del mar, y la altitud óptima es de 500 a 800 metros sobre nivel del mar" (p. 11). Además, la humedad relativa que necesita al año en promedio debe estar entre 70% y 80%.

Se concluye que la cosecha para la producción óptima del cacao se encuentra alrededor de la línea ecuatorial alrededor del mundo, por ello que destacan como los países que más producen los que se encuentran en Asia, América Latina, África y Oceanía.

La sobreproducción de grano de cacao a nivel mundial en 2017 ha tenido un impacto directo en los precios y ha disminuido un tercio su valor en el mercado aproximadamente, lo cual ha impactado en los productores (que en Perú en su mayoría son de baja escala, actividades familiares con 5 hectáreas de cultivo) y en el PBI de sus países, con millones de afectados indirectamente. Perú es el segundo productor de cacao de grano fino más grande del mundo y exporta el 90 % de la producción nacional. La ICCO (Organización internacional del cacao) recomienda coordinar estrategias entre países productores y consumidores de forma que se reduzca la crisis de los precios, ya que si continúan tomando decisiones descoordinadas se pronostica una crisis aún peor. Para llevar a cabo estas estrategias, es necesario estimar con mayor precisión la producción nacional de cacao.

El presente proyecto de trabajo de investigación surge de la necesidad de estudiar cómo el uso de técnicas Deep Learning y Long Short-Term Memory podrán ayudar a implementar un modelo de predicción de la producción de cacao. Existen muchos casos exitosos de estudios de modelos de predicción basados en estas técnicas; por ello se propone aplicarlo en el campo de la producción de cacao.

En este trabajo presenta la siguiente pregunta de investigación: ¿En qué medida un modelo de predicción ayuda en la estimación de la producción de cacao peruano, en el año 2021?; entonces, la importancia del presente trabajo de investigación es que nos ayudará a desarrollar un modelo de predicción de la producción de cacao para ayudar en la estimación de la producción de cacao peruano y reducir la crisis de precios.

## II. MARCO TEÓRICO

### A. Modelo de Predicción

Es una herramienta que usa datos y estadística para pronosticar resultados, el modelo se encarga de pronosticar la probabilidad de un resultado a partir de una porción fijada de datos.

Según Espino [3] los modelos de predicción surgen de la relación: rendimiento de una unidad de la muestra – de uno a más atributos conocidos de la unidad. Dicho modelo tiene como fin determinar la probabilidad de que exista un comportamiento específico cuando una unidad similar se encuentre en otra muestra diferente. Los modelos de predicción son utilizados mayormente

para ejecutar cálculos mientras se realizan operaciones, para evaluar el riesgo, de tal manera aporta para una mejor toma de decisiones.

Los modelos de predicción son importantes porque proveen información exacta y permite a los usuarios tomar precauciones, preparar los medios imprescindibles para minimizar las consecuencias de una acción, ya que es esencial tener información a detalle de los acontecimientos y resultados a largo plazo los cuales que reten nuestras presuposiciones. En el proyecto de investigación será esencial para determinar lo que puede llegar a pasar, no puede pronosticar lo que puede pasar más adelante, pero sí estudiar la información existente y precisar un resultado posible.

### B. Análisis predictivo

Es una manera de realizar un análisis estadístico que trata de extraer datos nuevos o históricos y usarla para pronosticar patrones de comportamiento. Esta técnica se puede usar a distintos acontecimientos ya sean del pasado, presente o futuro.

Según Vance [4] los análisis predictivos se utilizan para predecir un resultado determinado, utilizando datos históricos y herramientas analíticas como el aprendizaje automático y por supuesto, métodos estadísticos. El enfoque del análisis predictivo tiene como objetivo generar un grado significativo de precisión en los conocimientos futuros. Con la llegada de modelos y herramientas de análisis predictivo tecnológicamente avanzados, cualquier organización puede ahora utilizar datos pasados y actuales para predecir comportamientos y tendencias del negocio con precisión. Es esencial indicar que la eficacia y uso de los resultados va a depender de la forma como se analicen los datos.

### C. Deep Learning

Deep learning es una serie de algoritmos relacionados con las redes neuronales que tiene una mejor utilidad y un desempeño mayor que otros subcampos de machine learning. En vez de mostrarle al computador una relación grande de normas para solucionar un contratiempo, le entregamos un modelado que pueda examinar ejemplos y las acciones para adecuar cuando ocurran fallas.

En particular, la técnica Deep learning ha conseguido estos avances: Clasificación de imágenes, reconocimiento de voz y transcripción de escritura casi como un humano, mejora en la traducción automática,

conversión de texto a voz mejorada, los asistentes digitales también son producto de esta técnica, como Google Now y Amazon Alexa, una autonomía en la conducción a un nivel casi humano, entre otras [5].

#### D. Memoria a corto y largo plazo

LTSM es una arquitectura de red neuronal artificial recurrente que se usa en el campo deep learning y su uso ayuda a aumentar la calidad de las predicciones. Asimismo, son diseñadas específicamente para tratar datos secuenciales y dinámicos. Los LSTM son redes neuronales recurrentes (RNN) con una estructura ágil que intenta recordar las dependencias a largo plazo y prevenir los problemas habituales de RNN: gradientes en explosión y desaparición [6].

Las redes LSTM son convenientes en el trabajo de investigación para hacer pronósticos en base a datos de series de tiempo, ya que se pueden dar atrasos de duración incierta entre eventos importantes en una serie de tiempo. LSTM se desarrolla para solucionar la problemática que se puede dar trabajar con los RNN tradicionales.

#### E. Redes Neuronales Recurrentes

Es una arquitectura de Deep Learning, con una capacidad robusta para la clasificación y procesamiento de distintos datos en secuencia, por ejemplo: un texto, una imagen o video. Tiene como virtud la retroalimentación, ya que cada vez que ha analizado un conjunto de datos y como salida se obtiene una secuencia que permite el aprendizaje en memoria para las futuras entradas. Para ello es necesario el entrenamiento del Modelo de predicción, en él se indicará qué datos son los necesarios para obtener una proyección.

Oropeza [7] indica en su trabajo que es probable que las RNN necesiten gran cantidad de tiempo para el procesamiento de datos, ello dependerá del algoritmo que se utilice para entrenar las RNN. No obstante, a pesar de esto la velocidad con la que llegan al resultado esperado ignorando máximos y/o mínimos locales sobrepasa esta desventaja, en especial en redes que cuentan con un número reducido de neuronas. Es por ello por lo que en la actualidad las RNN son usadas como una de las principales herramientas para la predicción. Sin embargo, el generar una proyección toma más tiempo que otros tipos de Redes Neuronales

(feedforward y Elman), pero como resultado se obtiene una predicción confiable y con una tasa mínima de error.

#### F. Series de tiempo

Es una variable numérica presentada a lo largo del tiempo. Algunas de sus características son: periodicidad (cada cuanto se toman los datos), tendencia (cuando la serie crece o decrece a lo largo del tiempo), variabilidad (cuando la volatilidad varía o es constante a largo plazo). Se refiere también a datos estadísticos que se reúnen en intervalos de tiempo ya sean diarios, mensuales o anuales. Suele aplicarse a datos registrados en forma periódica como, por ejemplo: PIB, IPI o IBEX. Para Parra el estudio de las series temporales tiene como objetivo tener conocimiento del comportamiento que tendrá una variable en un período de tiempo para, a partir de dicho conocimiento, y en el supuesto de que no habrá cambios estructurales, poder realizar predicciones con el fin de determinar qué valor, en períodos de tiempo en el futuro, obtendrá la variable determinada inicialmente [8].

### III. METODOLOGÍA

La investigación presenta un diseño no experimental, esta herramienta se utiliza cuando la investigación se basa en estudiar un fenómeno en circunstancias normales y sus variaciones a lo largo del tiempo para luego hacer un análisis.

“Este tipo de investigación las variables independientes suceden y no son manipulables, no se puede influir en estas variables ni se tiene control directo sobre ellas, porque ya se produjeron, de la misma forma que las respectivas consecuencias” [9]. En el trabajo de investigación el diseño será no experimental transversal descriptivo, ya que tiene como finalidad tratar de llegar al conocimiento de poder predecir la producción de cacao y los valores que formulan una o más variables.

En este trabajo el tipo de investigación es Aplicada porque como resultado final se obtendrá una solución tecnológica, porque, “la investigación aplicada tiene como finalidad la solución de un problema, focalizado en la investigación y fortalecimiento de conocimientos aplicados, mejorando así el desarrollo científico y cultural” [10]. Entonces, la investigación aplicada tiene por finalidad solucionar contratiempos de una población u organizaciones, en nuestro trabajo se aborda la problemática de la producción de cacao peruano y como

solución tecnológica se obtendrá un modelo de predicción.

Por otro lado, se usará el enfoque cuantitativo para la recopilación de información para probar y comprobar hipótesis para luego examinar a detalle la información adquirida y sacar conclusiones.

Para Hernández [11] este enfoque se usa la recopilación y estudio de datos para contestar a las interrogantes y cerciorar que las suposiciones fijadas al principio del trabajo de investigación y se basa en la medición numérica, cálculo y usualmente en el empleo de estadística para fijar con precisión patrones de conducta de la población.

El enfoque cuantitativo es un proceso en el cual se hace uso de un razonamiento deductivo, cada fase lleva de manera racional a la que sigue, es útil para interpretar o pronosticar un suceso y determinar de manera exacta patrones de conducta de una población.

Se tomará como muestra 383 familias de la población según la fórmula para hallar el tamaño de la muestra, de esa manera formar la base de datos. En nuestro estudio tomaremos como conglomerados a los distintos departamentos donde se concentran la mayoría de las familias productoras de cacao.

#### A. Descripción del modelo de predicción

El modelo hace una estimación de la producción de cacao peruano. Para el desarrollo del modelo hay que tener en cuenta 3 elementos:

- Carga y procesamiento de datos se lleva a cabo por medio de un archivo de excel: que contiene datos de la producción de cacao por mes a nivel nacional.
- Elección de variables: Se determinan las variables que se usarán en el modelo. Las variables seleccionadas son cantidad de cacao kg por hectárea, costos de producción, tiempo de producción, entre otras.
- Ejecución del modelo: Aquí se lleva a cabo el entrenamiento del modelo el cual luego de varias pruebas para tener una mayor exactitud, se ejecuta y da como resultado la estimación de la producción de cacao.

#### B. Arquitectura del proyecto

La arquitectura del modelo tiene tres fases: Obtención de Data: Permite la carga y procesamiento por medio de un archivo de Excel; Red Neuronal: Una red neuronal que permiten generar predicciones; Modelo de predicción: Un modelo bajo TensorFlow previamente entrenado que responde a un conjunto de datos.

#### C. Planificación

##### Recolección de datos

Es esencial tener grandes cantidades de datos para el desarrollo del modelo, los cuales posibiliten asegurar excelentes resultados. Para el caso actual la data se encuentra en el sitio web de MINAGRI.

##### Guía de proceso

Entre las principales ventajas que tiene el usar PMBOK tenemos las siguientes:

Satisfacción en los clientes gracias al cronograma del proyecto se pueden ver las actividades a realizar.

Tiene un mayor control sobre los riesgos que pueden afectar al proyecto y permite prever para poder detectarlos a tiempo o disminuirlos.

Permite una mayor eficacia y por ello una mejor calidad en el entregable del proyecto a tiempo.

#### D. Ejecución

##### Base de datos

Como base de datos se cuenta con un archivo CSV donde se ubican los registros mensuales de la cantidad de producción de cacao desde el año 2007 al 2020. Para iniciar, tenemos que importar las librerías de Tensorflow y Keras, como se observa en la Figura N° 8, las que nos facilitarán la creación de un modelo de predicción usando Deep Learning y LSTM.

```
[1] import numpy as np
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM
```

Fig. 1 Librerías Tensorflow y Keras

Adicionalmente, debemos de importar las librerías de Pandas, para realizar la lectura de los registros mensuales que se encuentran en nuestro archivo csv, MinMaxScaler para la conversión de los valores en escala de 0 a 1 y Matplotlib.pyplot, para realizar gráficos con los datos.

```
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
```

Fig. 2 Importar librerías Pandas y Matplotlib.pyplot

Ya tenemos importadas las librerías necesarias en nuestro archivo Python, como siguiente paso debemos de realizar la lectura del archivo production.csv, indicamos que la columna 'Mes' será la inicial y la convertimos a formato de Date(fecha). Almacenamos la lectura en la variable dataset.

```
dataset = pd.read_csv('production.csv', index_col='Mes', parse_dates=['Mes'])
```

Fig. 3 Almacenar production.csv como un set de datos

Además, se aprecia que se está realizando la lectura correcta del archivo CSV llamando a nuestra variable dataset y la función 'head(170)', donde 170 es el total de registros mensuales que tenemos en nuestro archivo CSV. Luego, utilizamos matplotlib.pyplot para graficar nuestros datos, vamos a establecer fechas para indicar los meses que se encontrarán en el entrenamiento y en la validación del modelo de predicción.

```
set_entrenamiento = dataset['2008':'2017-12-01'].iloc[:,[True]]
set_validacion = dataset['2018-01-01':'2020'].iloc[:,[True]]

set_entrenamiento['Producción (miles de toneladas) - Cacao'].plot(legend=True)
set_validacion['Producción (miles de toneladas) - Cacao'].plot(legend=True)
plt.legend(['Entrenamiento (2008-2017)', 'Validación (2018-2020)'])
plt.xlabel('Tiempo')
plt.ylabel('Miles de toneladas')
plt.show()
```

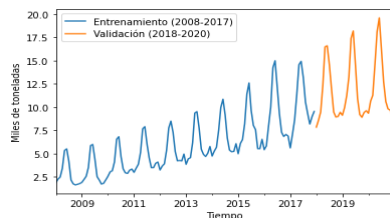


Fig. 4 Gráfico entrenamiento y validación

### Prototipo

Ahora convertimos los registros de la producción en valores de 0 a 1, utilizando la librería MinMaxScaler y se procede a llamar a nuestra variable set\_entrenamiento\_escalado para corroborar que se ha hecho la transformación a escala de 0 a 1.

```
escala = MinMaxScaler(feature_range=(0,1))
set_entrenamiento_escalado = escala.fit_transform(set_entrenamiento)
```

Fig. 5 Transformación de los datos de 0 a 1

La red LSTM tendrá como entrada "time\_step" que son la cantidad de datos consecutivos (Series temporales), y como salida un resultado (el pronóstico tomando los datos del "time\_step"). Se crea de tal manera el set de entrenamiento, con un intervalo de 10 datos y creamos los arreglos X\_train e Y\_train.

```
time_step = 3
X_train = []
Y_train = []
m = len(set_entrenamiento_escalado)
```

Fig. 6 Determinación de la cantidad de datos que tendrán las series temporales

Luego de ello, se realiza el llenado de los arreglos X\_train y Y\_train con un bucle for tomando como rango el tamaño del arreglo set\_entrenamiento\_escalado y el time\_step.

```
for i in range(time_step, m):
    X_train.append(set_entrenamiento_escalado[i-time_step:i,0])
    Y_train.append(set_entrenamiento_escalado[i,0])
X_train, Y_train = np.array(X_train), np.array(Y_train)
```

Fig. 7 Llenado de los arreglos X\_train e Y\_train Inicializamos el modelo en LSTM con las dimensiones anteriormente definidas. Colocamos mse en error para que el modelo nos muestre el error cuadrático medio, el cual consiste en comparar el valor real del set de entrenamiento con la predicción que realiza. Al ejecutar el modelo LSTM, se compilan las 100 épocas reduciendo cada vez el valor loss. De esta manera se puede esperar que el porcentaje de error de la predicción es mínimo.

```

Epoch 90/100
4/4 [=====] - 0s 9ms/step - loss: 0.0078
Epoch 91/100
4/4 [=====] - 0s 8ms/step - loss: 0.0078
Epoch 92/100
4/4 [=====] - 0s 8ms/step - loss: 0.0074
Epoch 93/100
4/4 [=====] - 0s 8ms/step - loss: 0.0086
Epoch 94/100
4/4 [=====] - 0s 8ms/step - loss: 0.0078
Epoch 95/100
4/4 [=====] - 0s 8ms/step - loss: 0.0076
Epoch 96/100
4/4 [=====] - 0s 9ms/step - loss: 0.0081
Epoch 97/100
4/4 [=====] - 0s 9ms/step - loss: 0.0071
Epoch 98/100
4/4 [=====] - 0s 9ms/step - loss: 0.0071
Epoch 99/100
4/4 [=====] - 0s 9ms/step - loss: 0.0071
Epoch 100/100
4/4 [=====] - 0s 8ms/step - loss: 0.0079
<tensorflow.python.keras.callbacks.History at 0x7fb5dd28c310>

```

Fig. 8 Ejecución del modelo LSTM

Posteriormente, debemos utilizar el set de validación para hacer el mismo proceso de transformar sus valores en una escala de 0 a 1 y colocarlos en un arreglo. Ajustamos a keras y se realiza la predicción con los valores de la validación con nuestro modelo ya entrenado y luego hacemos la transformación para que nos muestre los valores en cantidad de producción de cacao. Para ello, se crea la función para graficar las predicciones y los valores reales del set de validación.

```

def graficar_predicciones(real, prediccion):
    plt.plot(real[2:len(prediccion)],color='red', label='Valor real de Producción')
    plt.plot(prediccion, color='blue', label='Predicción de la Producción')
    plt.ylim(1.1 * np.min(prediccion)/2, 1.1 * np.max(prediccion))
    plt.xlabel('Tiempo')
    plt.ylabel('Miles de Toneladas')
    plt.legend()
    plt.show()

```

Fig. 9 Función para graficar la predicción – set validación

Obtenemos la siguiente gráfica, donde el color rojo es el valor real de producción y el azul de la predicción de la producción. De esta manera se comprueba el mínimo porcentaje de error que mostraba el modelo cuando se entrenaba.

```
graficar_predicciones(set_validacion.values,prediccion)
```

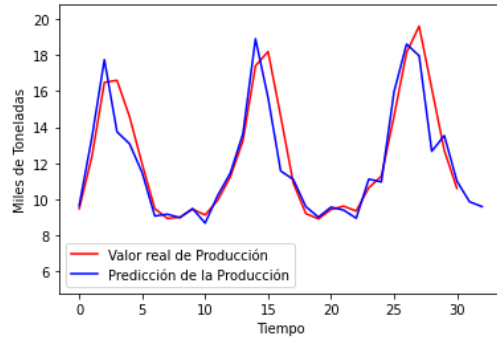


Fig. 10 Valor real - Predicción de la producción

#### IV. RESULTADOS

En este capítulo de los datos obtenidos del MINAGRI, se van a hacer las pruebas de hipótesis de la mano de los resultados de la estadística inferencial; además, se realiza la discusión de acuerdo con los antecedentes nacionales e internacionales.

##### A. Resultados de la estadística inferencial

Hi: El desarrollo de un modelo de predicción ayuda significativamente a estimar la producción de cacao peruano.

Ho: El desarrollo de un modelo de predicción no ayuda significativamente a estimar la producción de cacao peruano.

##### Correlaciones

		Modelo Predicción (Variable Independiente)	Producción de Cacao (Variable Dependiente)
Modelo Predicción (Variable Independiente)	Correlación de Pearson	1	,998**
	Sig. (bilateral)		,000
	N	37	37
Producción de Cacao (Variable Dependiente)	Correlación de Pearson	,998**	1
	Sig. (bilateral)	,000	
	N	37	37

Fig. 11 Resultados de correlación de Pearson

El valor de correlación de Pearson obtenido tiene un valor de 0.998 esto nos dice que existe una correlación positiva muy alta. Se determina que el modelo de predicción ayuda a estimar la producción de cacao peruano. Se acepta la hipótesis alterna y se rechaza la hipótesis nula.



### Hipótesis Específica 1

Hi: El modelo de predicción mejora notablemente el rendimiento de la producción de cacao peruano.

Ho: El modelo de predicción no mejora notablemente el rendimiento de la producción de cacao peruano.

		Modelo de Predicción (Variable Independiente)	Rendimiento (Dimensión)
Modelo de Predicción (Variable Independiente)	Correlación de Pearson	1	,996**
	Sig. (bilateral)		,000
	N	37	37
Rendimiento (Dimensión)	Correlación de Pearson	,996**	1
	Sig. (bilateral)	,000	
	N	37	37

Fig. 12 Estadística inferencial HE1

El valor de correlación R de Pearson obtenido tiene un valor de 0.996 esto nos dice que existe una correlación positiva muy alta. Se determina que el modelo de predicción mejora notablemente el rendimiento de la producción de cacao peruano. Se acepta la hipótesis alterna y se rechaza la hipótesis nula.

### Hipótesis Específica 2

Hi: El resultado del modelo de predicción cubre significativamente la demanda de la producción de cacao peruano.

Ho: El resultado del modelo de predicción no cubre significativamente la demanda de la producción de cacao peruano.

		Modelo de Predicción (Variable Independiente)	Demanda (Dimensión)
Modelo de Predicción (Variable Independiente)	Correlación de Pearson	1	,914**
	Sig. (bilateral)		,000
	N	37	37
Demanda (Dimensión)	Correlación de Pearson	,914**	1
	Sig. (bilateral)	,000	
	N	37	37

Fig. 13. Estadística inferencial HE2

El valor de correlación R de Pearson obtenido tiene un valor de 0.914 esto nos dice que existe una correlación positiva muy alta. Se determina que el modelo de predicción cubre significativamente la demanda de la producción de cacao peruano. Se acepta la hipótesis alterna y se rechaza la hipótesis nula.

### Hipótesis Específica 3

Hi: El modelo de predicción influye considerablemente en la oferta de la producción de cacao peruano.

Ho: El modelo de predicción no influye considerablemente en la oferta de la producción de cacao peruano.

		Modelo de Predicción (Variable Independiente)	Oferta (Dimensión)
Modelo de Predicción (Variable Independiente)	Correlación de Pearson	1	,960**
	Sig. (bilateral)		,000
	N	37	37
Oferta (Dimensión)	Correlación de Pearson	,960**	1
	Sig. (bilateral)	,000	
	N	37	37

Fig. 14 Estadística inferencial HE3

El valor de correlación de Pearson obtenido tiene un valor de 0.96 esto nos dice que existe una correlación positiva muy alta. Se determina que el modelo de predicción influye considerablemente en la oferta de la producción de cacao peruano. Se acepta la hipótesis alterna y se rechaza la hipótesis nula.

## V. DISCUSIÓN

El objetivo de este trabajo de investigación es desarrollar un modelo predictivo para la producción de cacao utilizando los procedimientos de Deep Learning y lstm, de tal manera que se pueda determinar la producción y reducir el riesgo de la devaluación del precio final.

Según Lazaro [12], asegura que es esencial para implementar un modelo de predicción para la producción el uso de herramientas y métodos estadísticos. Esto ayudó en el desarrollo del modelo de predicción para estimar la producción.

Sifuentes [13], implementa un modelo de predicción para aumentar el rendimiento académico de los alumnos en una universidad utilizando diferentes indicadores, para reducir el porcentaje de desaprobados. En contraste en la investigación se usaron distintos indicadores que ayuden a mejorar el rendimiento de producción de cacao.



Paz [14], desarrolla un modelo de redes neuronales artificiales y recurrentes que trabaje con deep learning para realizar pronósticos y comprueba que usando redes neuronales recurrentes se hacen mejores predicciones. Esto se tomó en cuenta en el trabajo y se optó por las redes neuronales recurrentes, ya que tiene un menor tiempo de procesamiento y son más precisas. Finalmente, se puede observar que los antecedentes previamente estudiados y aplicados en el proyecto, han garantizado lograr los resultados positivos que se buscaban con el modelo de predicción para poder estimar la producción de cacao usando las técnicas deep learning y lstm.

### CONCLUSIONES

Con los resultados obtenidos se puede concluir que el análisis estadístico respectivo de la información del sitio web de MINAGRI y al desarrollar el modelo propuesto ayuda a estimar la producción de cacao con un porcentaje de exactitud del 99.6% y reduce la crisis, también que, desarrollar un modelo de predicción mejora el rendimiento de la producción de cacao peruano.

Se determina que, el resultado del modelo de predicción cubre la demanda de la producción de cacao peruano y finalmente, se pudo demostrar la influencia del modelo de predicción con la oferta de la producción de cacao peruano.

### REFERENCIA

- [1]. Espino Timón, C. (2017) *Análisis Predictivo: Técnicas y modelos utilizados y Aplicaciones del Mismo - Herramientas open source que permiten Su Uso, Repositori Institucional (O2): Página de inicio*. Universitat Oberta de Catalunya. (2017)
- [2]. C. Oropeza. Modelado y simulación de un sistema de detección de intrusos utilizando redes neuronales recurrentes. Universidad de las Américas. Puebla, México.
- [3]. Coronati, A. , Andrade, JR , Bessa, RJ. "A deep learning method for forecasting residual market curves". (2020)
- [4]. Clercq, D. "Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data", Volumen 218, (pp. 390-399).
- [5]. Matich, D. "Redes Neuronales: Conceptos Básicos y Aplicaciones", (p. 8, 2001).
- [6]. Paz, D. "Herramientas Deep Learning para Predictive Analytics" (Tesis para maestría). Universidad de Concepción, Concepción, Chile, (2017)
- [7]. Lopez, D y Arco, L. Deep learning para la extracción de aspectos en opiniones textuales. Universidad de Oriente, Santiago de Cuba, Cuba. (2019)
- [8]. Fister, D. "Online Long Short-Term Memory Network for Stock Trading", (p. 5, 2019).
- [9]. Ordoñez E." Deep learning para la visión artificial e identificación del personal administrativo y docente de la universidad nacional micaela bastidas de Apurímac". Universidad Nacional del Altiplano, Puno, Perú. (2020).
- [10]. Hernandez, R, Fernandez, C, & Baptista L (2014). "Metodología de la investigación" (p. 152).
- [11]. Francois Chollet. "Deep Learning with Python", (p.11-12, 2017).
- [12]. Francisco Parra. "Estadística y Machine Learning con R", (p.208, 2019).
- [13]. *Estadísticas* (no date) *Producción de cacao se expandió 33,3% en enero de presente año*. Available at: <https://www.inei.gob.pe/prensa/noticias/produccion-de-cacao-se-expandio-333-en-enero-de-presente-ano-11468/>
- [14]. [14] Sherstinsky, A. (2020). "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network", (pp. 5-10, 2020).