



UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA ESTADÍSTICA E
INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E
INFORMÁTICA



MODELO UNIVARIANTE BOX JENKINS Y RED NEURONAL
PARA PRONOSTICAR EL NÚMERO DE GESTANTES
ADOLESCENTES, PUNO 2013-2023

TESIS

PRESENTADA POR:

CRISTIAN JOSE RODRÍGUEZ ROJAS

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2024



NOMBRE DEL TRABAJO

**MODELO UNIVARIANTE BOX JENKINS Y
RED NEURONAL PARA PRONOSTICAR E
L NUMERO DE GESTANTES ADOLESCEN
TES**

AUTOR

CRISTIAN JOSE RODRÍGUEZ ROJAS

RECuento de palabras

15503 Words

RECuento de caracteres

89521 Characters

RECuento de páginas

99 Pages

Tamaño del archivo

3.0MB

Fecha de entrega

Oct 12, 2024 6:47 PM GMT-5

Fecha del informe

Oct 12, 2024 6:48 PM GMT-5

● **13% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 11% Base de datos de Internet
- Base de datos de Crossref
- 7% Base de datos de trabajos entregados
- 2% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

● **Excluir del Reporte de Similitud**

- Material bibliográfico
- Material citado
- Bloques de texto excluidos manualmente
- Material citado
- Coincidencia baja (menos de 10 palabras)



Firmado digitalmente por CARPIO
VARGAS EDGAR ELOY
Motivo: Soy el autor del documento
Fecha: 12.10.2024 19:01:18 -05:00



Firmado digitalmente por JUAREZ
VARGAS Juan Carlos FAU
20145496170 soft
Motivo: Soy el autor del documento
Fecha: 14.10.2024 08:08:49 -05:00

Resumen



DEDICATORIA

Quiero dar las gracias a Dios, por su infinita bondad y por guiar mis pasos a través de este camino y a todos los miembros de mi familia. En especial a mis padres, Víctor Rodríguez Paquita y Melania Betty Rojas Lope que son las principales personas a las que debo mis logros, porque nunca me han abandonado, independientemente de las circunstancias. Les agradezco profundamente que me hayan enseñado a afrontar los retos con valentía y a mantener siempre la compostura ante las adversidades. Su presencia en mi vida ha sido un pilar fundamental en cada paso que he dado.

A mi hermana Yuliana, por su apoyo, su cariño y por ser una fuente inagotable de inspiración y motivación.

Cristian J. R. R.



AGRADECIMIENTOS

En primer lugar, a Dios, por darme la fortaleza, la sabiduría y la paciencia necesarias para enfrentar cada desafío en este camino académico.

A mis padres, Víctor y Betty, por su amor incondicional, su apoyo constante y sus enseñanzas. Gracias por ser mi refugio en los momentos difíciles y por inculcarme los valores y principios que guían mi vida. Su ejemplo de esfuerzo y dedicación ha sido mi mayor inspiración.

A mi hermana, por su cariño, comprensión y aliento en cada etapa de mi vida.

A mi asesor de tesis, Doc. Edgar Eloy Carpio Vargas por su invaluable guía, paciencia y consejos durante todo el proceso de investigación. Su conocimiento y experiencia han sido cruciales para la realización de este estudio.

Cristian J. R. R.



ÍNDICE GENERAL

	Pág.
DEDICATORIA	
AGRADECIMIENTOS	
ÍNDICE GENERAL	
ÍNDICE DE TABLAS	
ÍNDICE DE FIGURAS	
ÍNDICE DE ANEXOS	
ACRÓNIMOS	
RESUMEN	14
ABSTRACT.....	15
CAPÍTULO I	
INTRODUCCIÓN	
1.1. DESCRIPCIÓN DEL PROBLEMA.....	19
1.2. FORMULACIÓN DEL PROBLEMA.....	20
1.2.1. Problema general.....	20
1.3. OBJETIVOS DE LA INVESTIGACIÓN	20
1.3.1. Objetivo General	20
1.3.2. Objetivos Específicos.....	20
1.4. HIPÓTESIS DE LA INVESTIGACIÓN.....	21
1.4.1. Hipótesis General	21
1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN.....	21
1.6. LIMITACIONES DE LA INVESTIGACIÓN.....	22

CAPÍTULO II

REVISIÓN DE LITERATURA



2.1.	ANTECEDENTES DE LA INVESTIGACIÓN.....	23
2.2.	BASE TEÓRICA.....	27
2.2.1.	Técnicas de Predicción.....	27
2.2.2.	Serie de tiempo.....	27
2.2.3.	Características de una Serie de tiempo.....	28
2.2.3.1.	Estacionarias	28
2.2.3.2.	No estacionarias	28
2.2.4.	Componentes de una serie de tiempo.....	28
2.2.4.1.	Tendencia secular o regular	28
2.2.4.2.	Variación estacional o variación cíclica regular.....	29
2.2.4.3.	Variación cíclica irregular.	29
2.2.4.4.	Aleatoriedad.....	29
2.2.5.	Procesos estocásticos	29
2.2.6.	Tipos de procesos estocásticos.....	30
2.2.7.	Modelos autorregresivos (AR).....	32
2.2.8.	Modelos de medias móviles (MA).....	33
2.2.9.	Modelos ARMA.....	34
2.2.10.	Modelos ARIMA	35
2.2.11.	Modelos SARIMA	37
2.2.12.	Redes Neuronales Artificiales.....	37
2.2.13.	Red Neuronal Recurrente (RNN).....	38
2.2.14.	Criterios de evaluación y análisis de errores de pronósticos.....	42
2.2.15.	Error medio (ME).....	43
2.2.16.	Error cuadrático medio (RMSE)	44
2.2.17.	Error medio absoluto (MAE)	44



2.2.18. Porcentaje medio de error (MPE)	44
2.2.19. Error absoluto medio porcentual (MAPE)	45
2.3. DEFINICIÓN DE TÉRMINOS BÁSICOS.....	45
2.3.1. Serie	45
2.3.2. Modelo	45
2.3.3. Modelo Box – Jenkins.....	46
2.3.4. Estacionariedad	46
2.3.5. Estacionalidad	46
2.3.6. Modelo de pronostico.....	46
2.3.7. Lstm	46
2.3.8. Red neuronal	47
2.3.9. Entrenamiento de la red neuronal	47
2.3.10. Gestantes	47
2.3.11. Adolescente	47
2.3.12. Gestante adolescente	47
2.3.13. Embarazo adolescente	47
2.4. OPERACIONALIZACIÓN DE VARIABLES.....	48

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. POBLACIÓN	49
3.2. MUESTRA.....	49
3.3. MÉTODO DE RECOLECCIÓN DE DATOS	50
3.4. TÉCNICAS, INSTRUMENTOS, EQUIPOS Y MATERIALES.....	50
3.4.1. Técnicas de recolección	50
3.4.2. Técnicas estadísticas	50



3.5.	DISEÑO DE INVESTIGACIÓN	50
3.6.	TIPO DE INVESTIGACIÓN.....	50

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1.	APLICACIÓN BOX JENKINS Y UNA RED NEURONAL	51
4.2.	METODOLOGÍA BOX-JENKIS.....	53
4.2.1.	Identificación del Modelo	53
4.2.2.	Gráfico de tendencia, estacionalidad y residuos	55
4.2.3.	Función de Autocorrelación Simple y Parcial de la serie original.....	56
4.2.4.	Prueba Dickey-Fuller Aumentada para verificar la estacionariedad.....	57
4.2.5.	Serie diferenciada (Primera Diferencia).....	59
4.2.6.	Prueba Dickey-Fuller Aumentada para la serie diferenciada.....	60
4.2.7.	Gráfico de ACF y PACF de la serie diferenciada	61
4.2.8.	Estimación de modelos ARIMA	62
4.2.9.	Predicciones Box-Jenkins	67
4.2.10.	Validez del pronostico.....	70
4.3.	METODOLOGÍA RED NEURONAL RECURRENTE (RNN)	71
4.3.1.	Gráfico de la serie original de la serie.....	71
4.3.2.	Descomposiciones estacionales	72
4.3.3.	Conjuntos de datos de entrenamiento y prueba.....	74
4.3.4.	Escalando los datos	76
4.3.5.	Creación del modelo	78
4.3.6.	Transformación inversa pronósticos	80
4.3.7.	Predicciones futuras	83
4.3.8.	Métricas de evaluación del modelo.....	86



4.4. DISCUSIÓN	87
V. CONCLUSIONES	90
VI. RECOMENDACIONES	92
VII. REFERENCIAS BIBLIOGRÁFICAS.....	93
ANEXOS.....	97

ÁREA: Modelos predictivos.

TEMA: Modelo univariante Box Jenkins y Red Neuronal para pronosticar el número de gestantes adolescentes, puno 2013-2023

FECHA DE SUSTENTACIÓN: 24 de octubre del 2024



ÍNDICE DE TABLAS

	Pág.
Tabla 1 Operacionalización de variables	48
Tabla 2 Serie mensual de gestantes adolescentes Puno, Periodo 2013-2023	52
Tabla 3 Prueba de Dickey Fuller A. para verificar la estacionariedad	58
Tabla 4 Dickey – Fuller A. para verificar la estacionariedad - serie diferenciada....	60
Tabla 5 Selección de un modelo ARIMA óptimo.....	63
Tabla 6 Resultados del modelo ARIMA.....	64
Tabla 7 Coeficientes y Significancia del Modelo	65
Tabla 8 Pruebas de Diagnóstico del Modelo	66
Tabla 9 Predicciones estimadas del número de gestantes adolescentes 2024.....	67
Tabla 10 Predicciones estimadas del número de gestantes adolescentes 2025.....	68
Tabla 11 Métrica de evaluación del Modelo.....	70
Tabla 12 Data de entrenamiento	74
Tabla 13 Datos de prueba.....	75
Tabla 14 Datos de entrenamiento escalado.....	76
Tabla 15 Datos de validación escalados	77
Tabla 16 Modelo RNN.....	78
Tabla 17 Transformación inversa de pronósticos	80
Tabla 18 Comparación entre data test y predicción	81
Tabla 19 Predicciones futuras 2024.....	83
Tabla 20 Predicciones futuras 2025	84
Tabla 21 Métricas de evaluación del modelo.....	86



ÍNDICE DE FIGURAS

	Pág.
Figura 1 Ejemplos de ruido blanco	31
Figura 2 Estructura de una red neuronal recurrente	39
Figura 3 Estructura de una red neuronal recurrente	40
Figura 4 Mapa del departamento de Puno	49
Figura 5 Serie original de gestantes adolescente Puno periodo 2013-2023.....	53
Figura 6 Gráfico de la normalidad y distribución	54
Figura 7 Gráficos de los componentes: Tendencia, estacionalidad y residuos	55
Figura 8 Función de Autocorrelación Simple (ACF)	56
Figura 9 Función de Autocorrelación Parcial (PACF)	56
Figura 10 Serie temporal diferenciada.....	59
Figura 11 Autocorrelación (ACF) de la serie diferenciada.....	61
Figura 12 Autocorrelación parcial (PACF) de la serie diferenciada.....	62
Figura 13 Predicciones futuras 2024-2025	69
Figura 14 Serie original de gestantes adolescente	71
Figura 15 Descomposición estacional	72
Figura 16 Tendencia	72
Figura 17 Estacional	73
Figura 18 Residuos	73
Figura 19 Pérdida por época del Modelo	79
Figura 20 Comparación entre data test gestantes y predicciones	82
Figura 21 Predicciones de los años 2024 -2025.....	85



ÍNDICE DE ANEXO

	Pág.
ANEXO 1 Predicciones futuras 2026.....	97
ANEXO 2 Declaración jurada de autenticidad de tesis.....	98
ANEXO 3 Autorización para el depósito de tesis en el Repositorio Institucional.....	99



ACRÓNIMOS

MAE:	Error Absoluto Medio.
RMSE:	Raíz Cuadrada del Error Cuadrático Medio.
LSTM:	Memoria a Largo Plazo.
MSE:	Error Cuadrático Medio.
ARIMA:	Modelo Autorregresivo Integrado de Media Móvil.
RNN:	Red Neuronal Recurrente.
F.A.C:	Función de Autocorrelación.



RESUMEN

El objetivo de este estudio fue comparar dos enfoques diferentes, Box-Jenkins y Red Neural Recurrente (RNN), para determinar el modelo más adecuado para predecir la serie de gestantes adolescentes del departamento de Puno durante el periodo 2013-2023. El diseño de investigación empleado es no experimental y de tipo retrospectiva correlacional. La muestra de estudio estuvo compuesta por todo el conjunto de datos que comprendía 132 observaciones mensuales obtenidas del Repositorio Único Nacional de Información en Salud (Reunís). Después de realizar el análisis utilizando cada metodología, se generó pronósticos y los resultados se compararon con los datos reales. Esta comparación se realizó para determinar cuál de las dos metodologías producía predicciones más precisas. Se observó que la metodología Red Neuronal Recurrente arrojó un error de evaluación de pronóstico menor en comparación con la metodología Box Jenkins. Se utilizaron los indicadores MAE, RMSE, MPE, MAPE y el ME para evaluar la precisión del pronóstico. Como resultado, se considera que la metodología Red Neuronal Recurrente es la opción correcta para predecir resultados de la serie de gestantes adolescentes.

Palabras Clave: Box-Jenkins, Gestantes Adolescentes, Modelo Univariante, Red Neuronal Recurrente.



ABSTRACT

The objective of this study was to compare two different approaches, Box-Jenkins and Recurrent Neural Network (RNN), to determine the most appropriate model to predict the series of adolescent pregnant women in the department of Puno during the period 2013-2023. The research design used is non-experimental and retrospective correlational. The study sample was composed of the entire data set comprising 132 monthly observations obtained from the National Single Repository of Health Information (Reunís). After performing the analysis using each methodology, forecasts were generated and the results were compared with the real data. This comparison was made to determine which of the two methodologies produced more accurate predictions. It was observed that the Recurrent Neural Network methodology yielded a lower forecast evaluation error compared to the Box Jenkins methodology. The MAE, RMSE, MPE, MAPE and ME indicators were used to evaluate the forecast accuracy. As a result, the Recurrent Neural Network methodology is considered to be the correct option for predicting outcomes of the series of adolescent pregnant women.

Keywords: Box-Jenkins, Pregnant Adolescents, Univariate Model, Recurrent Neural Network.



CAPÍTULO I

INTRODUCCIÓN

La salud pública es un área de estudio y acción que se ocupa de proteger y mejorar la salud de la población, así como en la investigación de la prevención de enfermedades y lesiones. En este contexto, la gestación adolescente emerge como un problema crítico en el departamento de Puno, afectando no solo la salud y el bienestar de las jóvenes madres, sino también teniendo implicaciones significativas para sus familias.

La gestación adolescente es un fenómeno de preocupación global, y su impacto es particularmente pronunciado en regiones de bajos recursos. En estos contextos, las adolescentes enfrentan múltiples desafíos, como el acceso limitado a una educación sexual integral. Las causas de este problema son variadas y abarcan desde la falta de recursos económicos y las creencias religiosas y culturales, hasta la falta de información sobre métodos anticonceptivos y el miedo a buscar ayuda. Además, las normas sociales y las expectativas sobre el rol de género refuerzan este problema, dificultando que las adolescentes puedan tomar decisiones informadas y seguras sobre su salud sexual y reproductiva.

Los datos necesarios para llevar a cabo esta investigación fueron obtenidos del Repositorio Único Nacional de Información en Salud (REUNIS) del Ministerio de Salud, abarcando información mensual desde el año 2013 hasta el 2023.

(Fournies, 2015) expreso: “El método Box Jenkins fue generado en el año 1970 buscando facilitar el trabajo de los estadistas al construir un modelo de una serie temporal, para explicar su estructura y predecir la evolución de esta serie en el futuro. Una serie temporal se puede considerar como un conjunto de observaciones (datos), de una variable, tomados en intervalos regulares de tiempo. En particular, la metodología Box Jenkins es



un procedimiento de análisis estadístico para ajustar a una serie un tipo especial de modelos, denominados ARIMA” en su investigación denominada “Modelos ARIMA y Método Box Jenkins”.

“Las redes neuronales recurrentes (RNN) son redes que contienen bucles que almacenan información como una especie de memoria. Estas redes carecen de una estructura escalonada específica, lo que permite que las neuronas se enlacen arbitrariamente entre sí o consigo mismas. Las redes neuronales recurrentes se utilizan en inteligencia artificial y aprendizaje automático. Las redes neuronales deben aprender a lidiar con el complejo problema de los datos de series temporales utilizando una gran capacidad, estas redes tienen un estado oculto distribuido en el tiempo y pueden almacenar de manera eficiente una gran cantidad de información sobre el pasado.” (Quispe, 2023)

La investigación consta de cinco capítulos que a continuación se detallan:

Capítulo 1: Consideraciones generales, profundizamos en el proceso de normalización del tema, exploramos las motivaciones detrás de la realización del análisis, probamos la validez de nuestros análisis, reconocemos las limitaciones de la observación, discutimos las hipótesis y variables, y proporcionamos una explicación completa para el propósito de la tesis.

Capítulo 2: El marco teórico, sirve como base para nuestro estudio, donde examinamos el contexto histórico que rodea el tema, establecemos los principios teóricos que sustentan nuestra investigación y definimos términos clave. Además, proporcionamos referencias a otras tesis que comparten temas similares.

Capítulo 3: Profundizamos en los diversos aspectos de nuestra metodología de investigación, incluido el tipo de estudio realizado, el entorno del estudio, la población y



muestra involucradas, los métodos y herramientas utilizados para la recopilación de datos, así como la validez y confiabilidad de las herramientas. Además, proporcionamos una descripción general completa del diseño y detalles complejos de nuestros procedimientos de recopilación y procesamiento de datos.

Capítulo 4: Se presenta los resultados de nuestros esfuerzos de procesamiento de datos, mostrando los resultados obtenidos de nuestro análisis. Además, proporcionamos una discusión de antecedentes detallada que contextualiza e interpreta estos resultados, ofreciendo una comprensión más profunda de los hallazgos.

Capítulo 5: Se describen de forma exhaustiva las conclusiones del estudio, abarcando todos los objetivos planteados.



1.1. DESCRIPCIÓN DEL PROBLEMA

La gestación adolescente es un fenómeno que posee una incidencia significativa en el departamento de Puno, con casos registrados cada año. A pesar de la relevancia de este asunto, la falta de un análisis predictivo eficaz representa un vacío en la estrategia para abordar este problema. Un modelo predictivo confiable permitiría a los responsables de la toma de decisiones en salud pública identificar grupos de riesgo, optimizar la asignación de recursos y evaluar la efectividad de las intervenciones para prevenir la gestación adolescente.

Sin embargo, la incertidumbre sobre cuál es el modelo más adecuado para pronosticar el número de gestantes adolescentes genera una urgente necesidad de desarrollar modelos que permitan pronosticar la evolución del fenómeno y así tomar medidas oportunas.

En este contexto, la presente investigación se centra en determinar cuál de las dos metodologías, Box-Jenkins o Red Neuronal Recurrente (RNN), presenta mayor precisión y capacidad de generalización para la predicción del número de gestantes adolescentes en el departamento de Puno, considerando los datos disponibles.

Finalmente, el estudio aspira a proporcionar información valiosa para los responsables de la toma de decisiones en salud pública en el departamento de Puno, permitiéndoles implementar medidas preventivas y de intervención más efectivas y dirigidas, con el objetivo de reducir el número de gestantes adolescentes del departamento de Puno.



1.2. FORMULACIÓN DEL PROBLEMA

La formulación del problema se centra en determinar cuál modelo es más adecuado para pronosticar el número de gestantes adolescentes en el departamento de Puno. La metodología Box-Jenkins o una Red Neuronal Recurrente (RNN). Este análisis busca desarrollar estrategias efectivas de intervención y prevención en salud pública basadas en los hallazgos obtenidos.

1.2.1. Problema general

¿Cuál es el modelo más adecuado entre la metodología Box-Jenkins y una Red Neuronal Recurrente para pronosticar el número de gestantes adolescentes en el departamento de Puno durante el periodo 2013-2023?

1.3. OBJETIVOS DE LA INVESTIGACIÓN

1.3.1. Objetivo General

Determinar el modelo más adecuado entre la metodología Box-Jenkins y una Red Neuronal Recurrente, para el pronóstico del número de gestantes adolescentes en el departamento de Puno durante el periodo 2013-2023.

1.3.2. Objetivos Específicos

- Obtener un modelo univariante bajo metodología Box-Jenkins para el pronóstico del número de gestantes adolescentes en el departamento de Puno durante el periodo 2013-2023.
- Obtener un modelo de Red Neuronal para el pronóstico del número de gestantes adolescentes en el departamento de Puno durante el periodo 2013-2023.



- Evaluar las métricas de los modelos obtenidos para el pronóstico del número de gestantes adolescentes en el departamento de Puno durante el periodo 2013-2023.

1.4. HIPÓTESIS DE LA INVESTIGACIÓN

1.4.1. Hipótesis General

El modelo basado en una Red Neuronal Recurrente será más efectivo que el modelo Box-Jenkins para pronosticar el número de gestantes adolescentes en el departamento de Puno durante el periodo 2013-2023.

1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN

La gestación adolescente representa una preocupación significativa tanto a nivel social como de salud pública, debido a las implicaciones que conlleva para las gestantes adolescentes. En el contexto específico del departamento de Puno, es imperativo comprender y abordar este fenómeno de manera efectiva. Sin embargo, la identificación del modelo más idóneo para pronosticar el número de gestantes adolescentes en el departamento de Puno representa aún un desafío.

El propósito de esta investigación es cubrir este vacío de conocimiento al comparar dos enfoques de modelización: la metodología Box-Jenkins y una Red Neuronal Recurrente (RNN). A través de este estudio, se busca determinar cuál de estos modelos es más eficaz y preciso en el pronóstico del número de gestantes adolescentes en el departamento de Puno.

Con esta comprensión, podrán tomar decisiones informadas sobre estrategias y políticas para abordar la gestación adolescente en el departamento de Puno, utilizando los



hallazgos del estudio como una guía para mejorar las intervenciones y medidas preventivas.

Con el fin de contribuir a la investigación en el ámbito de la modelización de series temporales en salud pública, esta investigación es crucial debido a que abordará un tema importante y buscará proporcionar una solución efectiva para el pronóstico del número de gestantes adolescentes en Puno. Los resultados de esta investigación podrán ayudar a mejorar las medidas dirigidas a la prevención y atención de la gestación adolescente en el departamento de Puno.

1.6. LIMITACIONES DE LA INVESTIGACIÓN

Una de las limitaciones significativas de este estudio de investigación proviene de la fuente de datos empleada. La recopilación de datos alojados en el sitio web del Repositorio Único Nacional de Información en Salud (Reunís), estuvo en un Dashboard y no en una hoja de cálculo, lo cual para la extracción de la data fue complejo. Este proceso de obtención de datos se vio dificultado por la necesidad de la filtración de datos por meses y años para la construcción de la base de datos con el cual se tendría que llevar a cabo para dicha investigación, lo que conllevó un esfuerzo considerable.



CAPÍTULO II

REVISIÓN DE LITERATURA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

Leite et al., (2023). Indica una disminución del 8.4% en el número de nacimientos de madres adolescentes en 2020 comparado con 2019, con una tasa de fertilidad específica por edad de 44.1 por cada 1000 en adolescentes de 15 a 19 años. Estos resultados a pesar de los desafíos impuestos por la pandemia muestran una reducción notable en la natalidad adolescente, lo que se atribuye principalmente al distanciamiento social y la reducción de la exposición de las adolescentes a situaciones de riesgo, así mismo subraya la importancia de mantener servicios de salud accesibles durante la crisis sanitarias para continuar la tendencia de reducción de la natalidad adolescente, destacando una disminución significativa que podría influir en políticas públicas futuras.

Rostampour & Masoudi, (2023). Utilizando métodos de descomposición para explorar la estacionalidad y tendencias a largo plazo, aplicando el modelo Autorregresivo Integrado de Media Móvil (ARIMA) para ajustar una serie temporal univariante de incidencia de mordeduras de animales. Los resultados revelaron una tendencia estacional significativa y un aumento general en los incidentes de mordeduras de animales durante el período de estudio, Obteniendo un modelo ARIMA (2,0,0)(1,1,1). La existencia de una tendencia estacional en las mordeduras de animales también puede ayudar en la planificación de servicios de atención médica durante diferentes estaciones del año. Por lo tanto, el estudio resalta la importancia de implementar medidas proactivas para abordar el creciente problema de las mordeduras de animales en Irán.



Caira-Chuquineyra et al., (2023). En un estudio sobre la prevalencia del embarazo adolescente entre las adolescentes sexualmente activas en Perú, utilizando la Encuesta Demográfica y de Salud Familiar, indica que la prevalencia de embarazo adolescente fue del 30.9%, y se identificaron varios factores asociados como tener entre 17 y 19 años, estar casada o cohabitando, y pertenecer al grupo étnico Quechua. Los factores asociados con una menor prevalencia fueron estar empleadas, estar estudiando actualmente y tener conocimiento sobre el momento del ciclo en que puede ocurrir el embarazo.

Sardar et al., (2023). En un estudio sobre modelos de pronóstico automáticos basados en ARIMA y Prophet, junto con técnicas de aprendizaje automático. La metodología empleada incluyó la utilización de modelos de series temporales ARIMA y Prophet, y técnicas de aprendizaje automático como XGBoost, GLMNet y Random Forest, aplicadas a datos de COVID-19 de los países SAARC. Los resultados mostraron que el modelo ARIMA fue el más adecuado para predecir casos confirmados de COVID-19 en estos países, con una precisión significativamente superior a la de otros modelos. Por lo tanto, el estudio destacó la utilidad de combinar modelos estadísticos y de aprendizaje automático para mejorar la precisión en el pronóstico de enfermedades, proporcionando una herramienta valiosa para la planificación y respuesta de salud pública en situaciones de pandemia.

Quispe, (2023). Indica que la metodología empleada en su trabajo de investigación consistió en la recolección de datos históricos, la diferenciación de la serie para obtener estacionariedad y la identificación de modelos potenciales mediante el análisis de autocorrelación y autocorrelación parcial. Se estimó el modelo y se validó asegurando que los residuos se comportaran como ruido blanco. Los resultados mostraron que el modelo ARIMA proporcionaba una adecuada capacidad de predicción. Se concluyó que la metodología de Box-Jenkins era efectiva para describir y predecir el consumo de



energía, destacando la capacidad del modelo ARIMA para adaptarse y pronosticar con precisión bajo condiciones de estacionariedad y no estacionalidad.

Masini et al., (2021). Los resultados demostraron que los modelos híbridos y de ensamble mejoran significativamente la capacidad predictiva, con mejoras en la precisión predictiva en pruebas de capacidad predictiva superior. Del estudio resalta la eficacia de combinar distintos modelos y técnicas de aprendizaje automático para optimizar la predicción en economía y finanzas, marcando una mejora en la precisión predictiva general en un 5-10% comparado con métodos no combinados. Este avance subraya la importancia de los métodos híbridos en la predicción financiera de alta frecuencia, proporcionando un enfoque más robusto y adaptable para enfrentar los retos de los datos económicos modernos.

Kumar et al., (2020). Los resultados indicaron que el modelo LSTM predijo con éxito las tendencias futuras de casos COVID-19, con una posible conclusión de la pandemia hacia junio de 2020. Además, se compararon las tasas de transmisión con Italia y EE. UU., y se realizó una previsión para los siguientes 2, 4, 6, 8, 10, 12 y 14 días. La investigación concluyó que el análisis de series temporales mediante redes LSTM es efectivo para prever y planificar respuestas sanitarias, destacando la importancia de medidas proactivas para manejar brotes futuros. Esta metodología podría adaptarse para mejorar la precisión en la predicción y manejo de otras enfermedades infecciosas.

Sagheer & Kotb, (2019). Utilizando un enfoque de aprendizaje profundo, se implementó un modelo LSTM profundizado, optimizado mediante algoritmos genéticos, y se comparó su rendimiento con modelos estadísticos y de redes neuronales menos complejos. Además reveló que el modelo DLSTM mejoró la precisión del pronóstico en comparación con los métodos tradicionales, demostrando una mejora significativa en la



precisión de las predicciones, con una disminución en el error de pronóstico medido por el RMSPE (Root Mean Square Percentage Error) del 3.4% al 2.9%. Por lo tanto las arquitecturas profundas como DLSTM son más eficaces para capturar la complejidad y la no linealidad de los datos de series temporales.

Calla, (2019). Indica que el modelo univariante que permitiera describir y predecir el consumo mensual de bolsas de cemento, utilizando la metodología de Box-Jenkins. La investigación aplicó el modelo ARIMA (0,2,1)(0,1,1) tras la identificación, estimación y validación mediante análisis de residuos, asegurando que estos fueran compatibles con un ruido blanco. Por lo tanto indica que el modelo univariante era el más adecuado para pronosticar las ventas, proporcionando una herramienta efectiva para la planificación empresarial y reduciendo el riesgo de sobreproducción y subproducción.

Zambrano et al., (2019). En un estudio del nivel de conocimiento sobre el papilomavirus humano (HPV) entre adolescentes embarazadas en un hospital de São Paulo, Brasil. Utilizaron una metodología descriptiva para recopilar datos de junio a diciembre de 2017, aplicando encuestas a adolescentes embarazadas menores de 18 años. Revelaron que, aunque el 80.92% había oído hablar del HPV, solo el 50.66% había recibido información en la escuela, y un 66.45% no sabía cómo se transmitía el virus. Además, la investigación destacó que más de la mitad no estaba vacunada y muchas no reconocían la conexión entre el HPV y el cáncer cervical. Indica que hay una brecha significativa en la prevención y conocimiento del HPV, lo que indica la necesidad urgente de mejorar la educación y las intervenciones de salud pública para esta población vulnerable.



Ayala & Samillán, (2019). En un estudio sobre comparación de dos metodologías, Box Jenkins y Redes Neuronales Artificiales, para pronosticar el precio de cierre mensual de las acciones del Banco de Crédito del Perú en la Bolsa de Valores de Lima. Encontró que las Redes Neuronales Artificiales superaron a Box Jenkins en la precisión de pronóstico, con menores errores de evaluación según indicadores como MAE, RMSE, MPE, MAPE y ME. La selección de la red neuronal se basó en el criterio de parsimonia, lo que sugiere que las Redes Neuronales Artificiales son una mejor opción metodológica para pronosticar series financieras.

2.2. BASE TEÓRICA

2.2.1. Técnicas de Predicción

En el análisis de series de tiempo, es esencial diferenciar entre pronósticos dentro de muestra, que se basan en los datos usados para ajustar el modelo, y fuera de muestra, que se aplican a datos nuevos. Además, los pronósticos pueden ser estáticos, limitados a una proyección para el siguiente periodo utilizando la última información disponible, o dinámicos, que utilizan el resultado de un pronóstico anterior para realizar proyecciones futuras, útiles en escenarios de cambios rápidos

2.2.2. Serie de tiempo

“Las series de tiempo o serie cronológico es un conjunto de datos observados en forma secuencial, generalmente en intervalos de tiempos iguales” (Córdova, 2006, p.355).



2.2.3. Características de una Serie de tiempo

2.2.3.1. Estacionarias

El concepto de estacionariedad gira en torno a la noción de que las leyes que gobiernan el comportamiento de un proceso permanecen constantes a lo largo del tiempo. Esencialmente, el proceso alcanza un estado de equilibrio estadístico. Más específicamente, un proceso $\{Y_t\}$ se considera estrictamente estacionario cuando la distribución conjunta permanece sin cambios para cualquier combinación de puntos de tiempo (t_1, t_2, \dots, t_n) y tiempo de retraso (k) . (Cryer & Chan, 2008).

2.2.3.2. No estacionarias

Estas series muestran cambios en las tendencias y/o fluctuaciones a lo largo del tiempo. Las alteraciones en los valores promedio indican una inclinación a largo plazo hacia el crecimiento o la disminución, lo que hace que la serie no gire en torno a un valor fijo. (Cryer & Chan, 2008).

2.2.4. Componentes de una serie de tiempo

2.2.4.1. Tendencia secular o regular

“Representa el comportamiento predominante de la serie. Esta puede ser definida vagamente como el cambio de la media a lo largo de un extenso periodo de tiempo”. (Rios, 2008).



2.2.4.2. Variación estacional o variación cíclica regular.

“Es un movimiento periódico que se producen dentro de un periodo corto y conocido. Este componente está determinado, por ejemplo, por factores institucionales y climáticos”. (Rios, 2008).

2.2.4.3. Variación cíclica irregular.

“Caracterizado por oscilaciones alrededor de la tendencia con una larga duración, y sus factores no son claros. Por ejemplo, fenómenos climáticos, que tienen ciclos que duran varios años”. (Rios, 2008).

2.2.4.4. Aleatoriedad

Estos movimientos, caracterizados por la falta de un patrón discernible, pueden atribuirse a multitud de factores. Su naturaleza es altamente impredecible y abarca una amplia gama de movimientos dentro de una serie temporal que no pueden clasificarse como tendencias, variaciones estacionales o acciones cíclicas. (Rios, 2008).

2.2.5. Procesos estocásticos

Una secuencia de variables aleatorias $\{Y_t: t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ se denomina proceso estocástico y se utiliza como modelo para una serie de tiempo observada. Se sabe que la estructura de probabilidad completa del proceso está determinada por el conjunto de distribuciones sobre todos los conjuntos finitos de Y 's. (Cryer & Chan, 2008).

2.2.6. Tipos de procesos estocásticos

- Proceso estocástico estacionario

Si un proceso estocástico permanece estacionario, significa que su promedio y variabilidad se mantienen constantes en el tiempo. Además, la covarianza entre dos períodos de tiempo está determinada únicamente por el desfase o brecha entre ellos, más que por el momento específico de esos períodos. (Villavicencio, s.f.)

Sea X_t una serie de tiempo entonces con estas propiedades:

$$\text{Media } E(X_t) = E(X_{t+k}) = \mu$$

$$\text{Varianza } V(X_t) = V(X_{t+k}) = \sigma^2$$

$$\text{Covarianza } Y_k = E [(X_t - \mu) (X_{t+k} - \mu)]$$

“En resumen, si una serie de tiempo es estacionaria, su media, su varianza y su autocovarianza (en diferentes rezagos) permanecen iguales sin importar el momento en el cual se midan; es decir, son invariantes respecto al tiempo”. (Villavicencio, s.f.).

- Ruido blanco

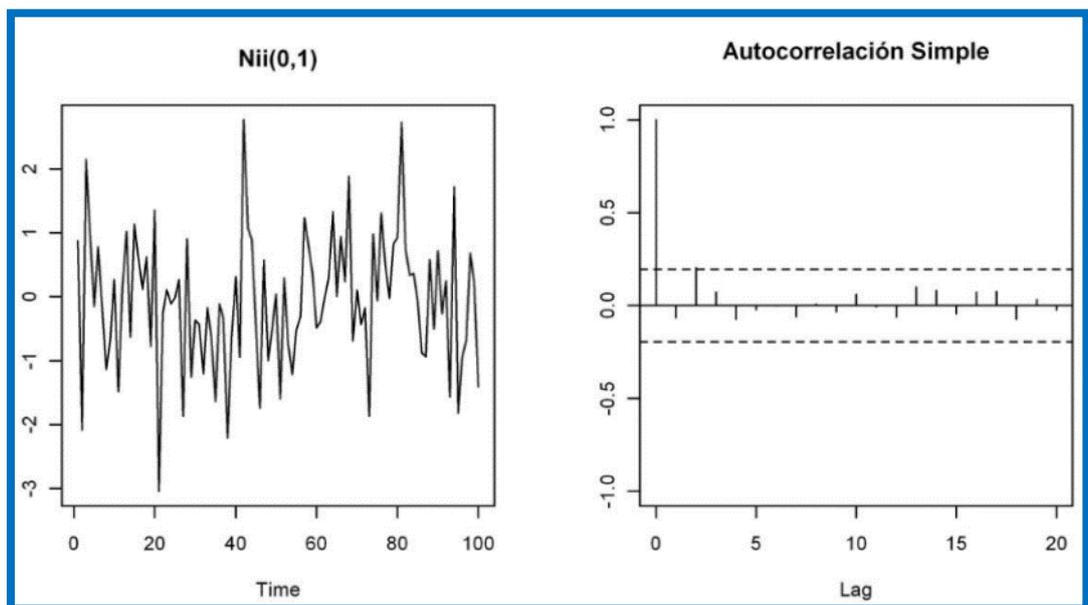
El ruido blanco, como proceso estocástico, es una señal aleatoria que muestra una falta de correlación estadística entre los valores de su señal en diferentes momentos. Esto significa que los valores de la señal a lo largo del eje del tiempo no están relacionados entre sí. El ruido blanco encuentra aplicación como término de error. (Fournies, 2015).

El proceso de ruido blanco, a menudo citado como un ejemplo significativo de proceso estacionario, consiste en una secuencia de variables aleatorias $\{\varepsilon_t\}$ que son independientes y están distribuidas de manera idéntica. Su importancia no surge de su naturaleza inherente como proceso estacionario. No sólo es un modelo intrigante en sí mismo, sino que también sirve como base para construir numerosos procesos valiosos utilizando ruido blanco. El término "ruido blanco" se deriva de la observación de que, al igual que la luz blanca, todas las frecuencias están igualmente presentes en el modelo cuando se analiza. Normalmente, asumimos que el proceso de ruido blanco tiene un valor de cero y representa la varianza de (ε_t) . (Cryer & Chan, 2008)

$$\varepsilon_t \sim N(0, \sigma^2) \text{Cov}(\varepsilon_{ti}, \varepsilon_{tj}) = 0 \forall ti \neq tj$$

Figura 1

Ejemplos de ruido blanco



NOTA: (Villavicencio, s.f.)

- Camino aleatorio

El proceso de paseo aleatorio simple sirve como modelo útil, proporcionando una aproximación razonable, para una amplia gama de fenómenos, incluida la fluctuación de los precios de las acciones y el comportamiento de pequeñas partículas suspendidas en un fluido conocido como movimiento browniano. (Cryer & Chan, 2008).

2.2.7. Modelos autorregresivos (AR)

“Los procesos autorregresivos son como su nombre indica: regresiones sobre sí mismos. Específicamente, un proceso autorregresivo de orden p $\{Y_t\}$ satisface la ecuación.”(Cryer & Chan, 2008)

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t$$

“Comenzamos considerando un proceso autorregresivo simple de primer orden. La corriente el valor del período de $\{y_t\}$ se explica por su anterior, una constante c , y un proceso de error $\{\varepsilon_t\}$ ” (Pfaff, 2008).

$$Y_t = c + \varphi_1 Y_{t-1} + \varepsilon_t$$

En esencia, este proceso puede describirse como un fenómeno de ruido blanco. Puede entenderse como una ecuación en diferencias de primer orden, con variaciones dependiendo del valor de φ . Si el valor absoluto de φ es mayor o igual a 1, el proceso no permanecerá constante en el tiempo y, como resultado, se considera no estacionario.

Los modelos autorregresivos se indican con el acrónimo AR, seguido del orden del modelo, como AR (1) o AR (2). El orden del modelo representa el

número de observaciones previas de la serie temporal que se incluyen en la ecuación. La forma general de un modelo autorregresivo, AR (p), se puede expresar de la siguiente manera:(Melo Mayta, 2016).

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Donde ε_t es una variable de “ruido blanco”, siempre y cuando cumpla lo siguiente:

- Media nula
- Varianza constante
- Covarianza nula entre errores correspondientes a observaciones diferentes.

2.2.8. Modelos de medias móviles (MA)

“En el caso donde solo un número finito de los pesos no es cero, tenemos lo que es llamado proceso de promedio móvil. En este caso, cambiamos la notación un tanto y escribimos.” (Cryer & Chan, 2008).

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q}$$

El nombre dado a esta serie es media móvil de orden q, que comúnmente se abrevia como MA(q). El término "media móvil" se utiliza porque Y_t se calcula aplicando ponderaciones de $1, -\theta_1, -\theta_2, \dots, -\theta_q$ a las variables $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ y luego cambiando el peso. Para obtener Y_{t+1} y valores posteriores, el proceso consiste en aplicarlos a $\varepsilon_{t+1}, \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q+1}$.(Cryer & Chan, 2008).

En esta sección exploraremos el concepto de modelar un proceso como un promedio móvil finito de sus shocks. Este tipo de proceso se conoce como MA(q), donde el parámetro q representa el retardo máximo de los shocks que se consideran en el modelo. Para comenzar nuestro análisis, examinaremos un proceso MA (1) representado por la ecuación $y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$, como se muestra en la ecuación (1.21). (Pfaff, 2008).

Los indicadores conocidos como medias móviles tienen la capacidad de reducir o suavizar las fluctuaciones, independientemente de su duración, ya sea a corto, medio o largo plazo. Estos modelos se denominan comúnmente MA, seguido del orden entre paréntesis, similar a los modelos autorregresivos. La notación MA(q) significa un modelo de media móvil con un orden de q. Es importante señalar que el proceso de media móvil implica combinar variables de ruido blanco utilizando coeficientes "theta" como ponderaciones para la combinación lineal. (Melo Mayta, 2016).

2.2.9. Modelos ARMA

El comportamiento de los estimadores de parámetros a medida que se acercan al infinito proporciona información valiosa sobre los desafíos asociados con el ajuste de los modelos ARMA a los datos. Para ilustrar, consideremos una serie de tiempo que sigue un proceso AR(1) e intentamos ajustar un modelo AR(2) a los datos. ¿Hay algún problema potencial con este enfoque? En un sentido más amplio, es importante comprender por qué son importantes estas consideraciones. ¿No es prudente utilizar modelos AR expansivos para capturar con precisión las complejidades del proceso? Es lógico que si el proceso es genuinamente un AR(1), la importancia de los otros parámetros autorregresivos sería insignificante.

La respuesta afirmativa a esta pregunta es inequívoca. Cuando superamos el umbral de t , nuestras estimaciones de parámetros se vuelven menos exactas o menos precisas. Esto es evidente al comparar un proceso AR(1) con otro proceso AR(1) para n grande. Sin embargo, si comparamos un proceso AR(2) con un proceso AR(1) para n grande, la sobreestimación ocurre porque $2 = 0$. Es importante señalar que esta sobreestimación puede servir como una valiosa herramienta de diagnóstico. Para ilustrar, consideremos una situación en la que tenemos un conjunto de datos y determinamos que un modelo AR(2) representa adecuadamente los datos. Si decidimos agregar un parámetro adicional y ajustar un modelo AR(3), esperaríamos que el modelo resultante fuera muy similar al modelo AR(2) original. Al igual que en el caso del modelo AR(p), el modelo MA(q) se puede expresar como:

$$Y_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) w_t. \text{ (Shumway \& Stoffer, 2015)}$$

Como el proceso de media móvil autorregresiva mixta con órdenes p y q , al que nos referiremos como ARMA (p, q). Ahora, comencemos examinando un escenario específico significativo. (Cryer & Chan, 2008)

2.2.10. Modelos ARIMA

Si observamos una serie de tiempo $\{Y_t\}$ y encontramos que sigue una media móvil autorregresiva integrada, podemos determinar esto examinando la diferencia d th $W_t = \nabla^d Y_t$, que debería ser un proceso ARMA estacionario. En el caso de que $\{W_t\}$ siga un modelo ARMA (p, q), podemos concluir que $\{Y_t\}$ es un proceso ARIMA (p, d, q).

En la mayoría de los casos, es conveniente suponer que $d = 1$ o posiblemente 2 para fines prácticos. Examinemos un proceso ARIMA $(p, 1, q)$ usando la variable W_t , que representa la diferencia entre $Y_t - Y_{t-1}$, Esto se puede expresar como: (Cryer & Chan, 2008)

$$W_t = \varphi_1 W_{t-1} + \varphi_2 W_{t-2} + \dots + \varphi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Para comenzar, la serie se examina visualmente para determinar si presenta estacionariedad. Si existen dudas sobre el cumplimiento de esta condición, el investigador debe transformar adecuadamente la serie antes de continuar. Esto puede implicar eliminar una tendencia determinista o tomar las primeras diferencias a lo largo del tiempo. Además, si hay inestabilidad de la varianza con mayores fluctuaciones en el tiempo, el uso de los valores logarítmicos de la serie puede solucionar este problema. Después de especificar un modelo ARMA (p, q) , el siguiente paso es estimar un modelo preliminar. El principio ML se utiliza para distinguir entre diferentes especificaciones de modelos mediante el cálculo de criterios de información y/o la realización de pruebas de índice de verosimilitud. Estas herramientas proporcionan un medio adicional para determinar un orden de retraso adecuado para los modelos ARMA (p, q) , además de la decisión de orden derivada de ACF e PACF. Un requisito previo para utilizar modelos ARIMA es que la serie temporal debe ser estacionaria. En consecuencia, si se comienza con una serie temporal no estacionaria. Para lograr una serie temporal estacionaria, el paso inicial es realizar el proceso de diferenciación. Si es necesario diferenciar la serie temporal "d" veces para lograr una serie estacionaria, entonces se aplica un modelo ARIMA (p, d, q) , donde "d" representa el orden de diferenciación utilizado. (Coghlan, 2014)

2.2.11. Modelos SARIMA

Los patrones estacionales son una ocurrencia común en los datos de series temporales, donde un componente específico se repite en todas las observaciones. Por ejemplo, los datos mensuales suelen exhibir un ciclo estacional de 12 (que representa los 12 meses de un año), mientras que los datos trimestrales siguen un ciclo de 4 (que representan los 4 trimestres de un año). Para abordar esta estacionalidad, los investigadores han desarrollado versiones generalizadas de los procesos ARIMA conocidas como modelos **SARIMA** (Modelo de media móvil integrada autorregresiva estacional). (Gekko, 2017).

La ecuación:

$$\Phi(B) \Delta^d X_t = \theta(B) \alpha_t$$

Utilizamos la notación $X_t \sim \text{ARIMA}(p, d, q) \times (P, D, Q)$ s para representar nuestros modelos SARIMA. Estos modelos son esencialmente modelos ARIMA (p, d, q) , pero con el componente agregado de los residuos α_t que siguen una estructura ARIMA (P, D, Q) . En el caso de ARIMA (P, D, Q) , proponemos modelos ARIMA que incorporan operadores definidos en B s y sus potencias sucesivas. El concepto de regiones permitidas en los modelos SARIMA es análogo al de regiones permitidas en los procesos ARIMA, con la diferencia de que se expresan únicamente en términos de potencias de B s. (Gekko, 2017).

2.2.12. Redes Neuronales Artificiales

Definición

Las arquitecturas de procesamiento paralelo conocidas como RNA proporcionan mecanismos innovadores que se pueden aplicar a una amplia gama de problemas, permitiendo la resolución de desafíos que son arduos o



inalcanzables utilizando enfoques convencionales. La utilización de RNA se ha ampliado para abarcar predicción, clasificación, reconocimiento y otros problemas que son inherentemente imprecisos y para los cuales los métodos de programación tradicionales ofrecen soluciones restringidas o insuficientes. (Pérez Delgado & Quintín Martín, 2003).

2.2.13. Red Neuronal Recurrente (RNN)

LA RNN poseen estados ocultos que se extienden a lo largo del tiempo, lo que les permite retener de manera efectiva una cantidad sustancial de información histórica. Similar a una red neuronal direct-feed al incorporar nodos que introducen dinámicas no lineales, la alimentación directa permite que las redes neuronales capturen de manera efectiva la dinámica intrincada de los datos de series temporales. Las redes neuronales se pueden clasificar en dos tipos: dinámicas y estáticas. Las redes neuronales estáticas generan resultados directamente a partir de los datos de entrada a través de conexiones de retroalimentación, con la información fluyendo en una sola dirección. Estas redes carecen de elementos de retroalimentación. Por otro lado, las redes neuronales dinámicas producen resultados que dependen de la entrada actual, así como de las entradas, salidas y/o estados ocultos de la red. (Lewis, 2017)

La red pasa por el siguiente proceso en cada paso de tiempo:

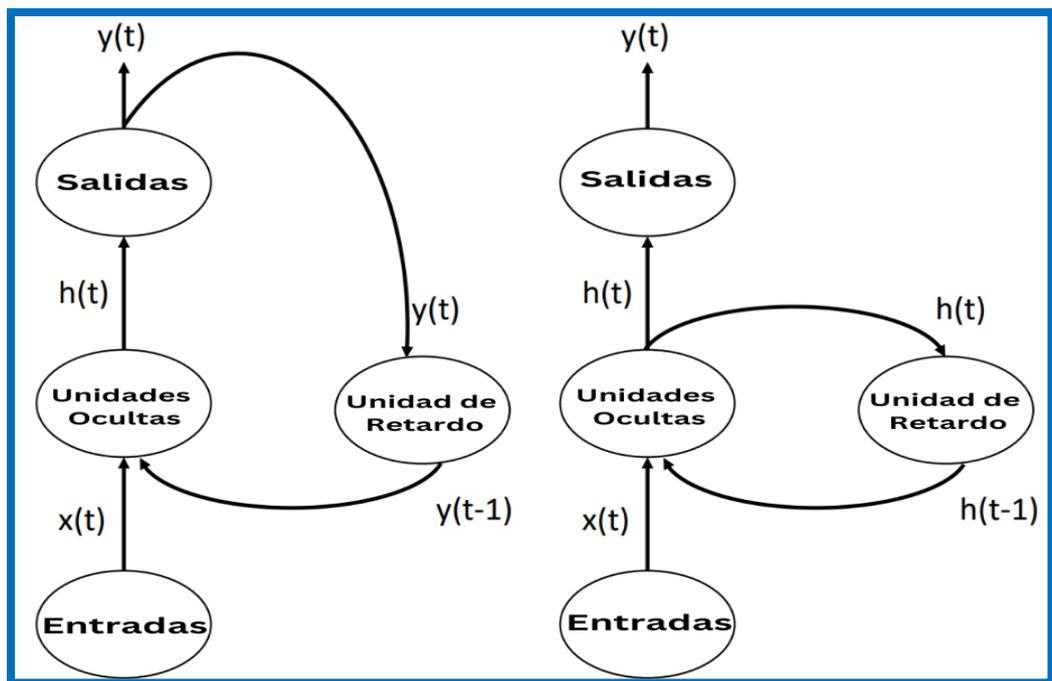
1. Se procesan los atributos de entrada (x_t).
2. El estado oculto se modifica empleando funciones de activación (h_t).
3. Utilizando esta información, puede pronosticar el resultado de su producción (y_t) (Lewis, 2017)

El proceso descrito aquí se parece a la red neuronal de retroalimentación (Lewis, 2017).

Sin embargo, la información almacenada en el retraso se reintroduce en las unidades ocultas como entrada complementaria (Lewis, 2017).

Figura 2

Estructura de una red neuronal recurrente



Nota: Neuronal Networks for Time Series Forecasting with R – Lewis, 2017.

Se ha descubierto que, con una cantidad suficiente de neuronas y tiempo, los RNN tienen la capacidad de realizar cualquier cálculo que pueda realizar una computadora. Este fenómeno se conoce en el campo de la informática como lograr la finalización total. Esta capacidad permite a los RNN resolver eficazmente una amplia gama de problemas. Lamentablemente, la aplicación de soluciones teóricas a problemas informáticos a menudo se ve obstaculizada por limitaciones prácticas como la memoria finita y las limitaciones de tiempo. (Lewis, 2017).

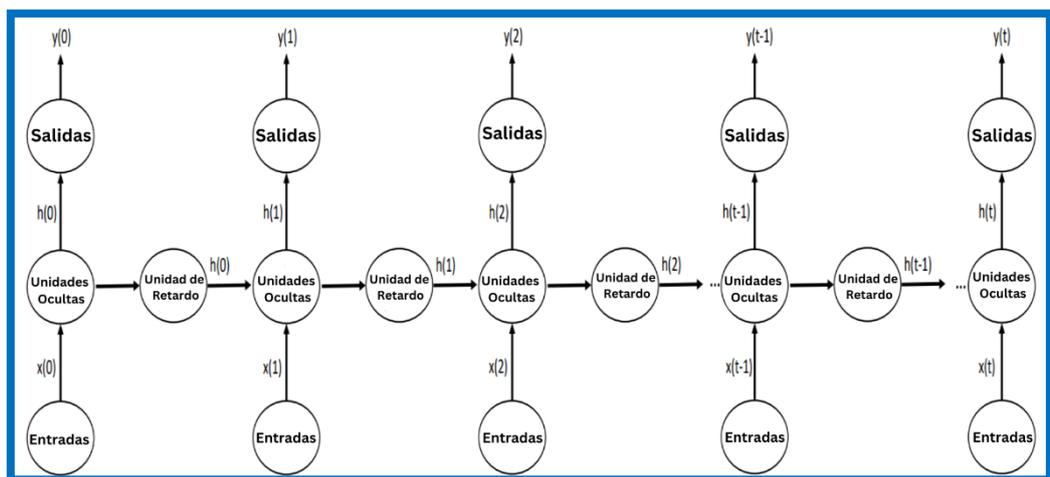
Al incorporar una unidad de retardo, la red gana la capacidad de retener información en su memoria a corto plazo. Esto se logra almacenando los valores de activación de la capa oculta y/o los valores de salida del paso de tiempo anterior. Estos valores almacenados luego se reintroducen en la red durante el siguiente paso de tiempo. Esencialmente, el RNN posee una "memoria" que captura los cálculos realizados por las unidades ocultas en un paso de tiempo anterior. (Lewis, 2017).

- **Backpropagation**

El algoritmo de retropropagación se puede aplicar para entrenar cualquier red neuronal de alimentación hacia adelante realizando algunas modificaciones. Si bien la idea de mostrar el RNN a lo largo del tiempo puede parecer compleja, el concepto detrás de ella es bastante sencillo. (Lewis, 2017).

Figura 3

Estructura de una red neuronal recurrente



Nota: Neuronal Networks for Time Series Forecasting with R – Lewis, 2017.

Cuando nos referimos a "implementación", esencialmente estamos describiendo el proceso de construcción de una red que abarca todos los pasos de tiempo que se analizan. El RNN implementado es esencialmente una red neuronal



profunda, en la que cada capa representa un paso de tiempo específico en el RNN original. (Lewis, 2017).

Para mejorar la efectividad del modelo y minimizar el error de generalización, es crucial determinar la cantidad de semillas, que representa el conjunto de números aleatorios utilizados para ajustar los pesos sinápticos en las capas ocultas y las salidas durante cada Época. Además, la cantidad de pruebas realizadas determinará la frecuencia con la que se evalúa y mejora la coherencia del modelo.

- **Metodología paso hacia adelante:**

Los pesos de un modelo de tipo neuronal a menudo se pueden mejorar reajustándolos cuando se obtienen nuevos datos. En nuestro estudio, necesitaremos generar seis predicciones con un mes de anticipación, lo que significa recalibrar el modelo seis veces. Esto se puede lograr mediante un sencillo proceso "for". (Lewis, 2017).

Empecemos configurando los parámetros principales:

Darse cuenta de:

Dentro del inicio del parámetro se encuentra el valor final de la fila correspondiente al mes final de la muestra de entrenamiento. En consecuencia, adopta este valor último.

Dentro del último parámetro, hay un conteo de 6 para los pronósticos de avance hacia adelante, específicamente un paso a la vez.

La variable de contador, k , se utiliza para incrementar el atributo de observaciones en un solo paso de tiempo (mes). (Lewis, 2017).



Vista general del bucle;

1. Abra el bucle “for”
2. Preparar los datos de entrenamiento.
3. Especifique el modelo.
4. Preparar los datos de prueba.
5. Hacer predicciones y cierra el bucle “for” (Lewis, 2017).

Normalmente, los pronósticos un paso adelante son más precisos al predecir los precios reales debido a su proximidad a los datos actuales. Estos pronósticos utilizan las observaciones y datos más recientes, lo que resulta en una mayor precisión. Sin embargo, es importante señalar que estos pronósticos requieren un tiempo de procesamiento más largo ya que es necesario volver a estimar el modelo para cada nueva observación. (Lewis, 2017).

2.2.14. Criterios de evaluación y análisis de errores de pronósticos

GestioPolis.com (2002), argumento sobre la evaluación de pronósticos, “Implica comparar los valores del pronóstico con los valores históricos reales. Frecuentemente, el examen de los patrones de errores lleva al analista a modificar el procedimiento de pronósticos”.

INGENIO EMPRESA, (2016), argumento sobre el cálculo del error de pronóstico, “Su cálculo nos permite tomar decisiones frente a qué método de pronóstico es el mejor y logran detectar cuando algo en nuestra previsión de la demanda no está marchando bien, con lo que conseguimos cambiar el rumbo de nuestras decisiones a fin tomar las mejores elecciones”.



Las predicciones se basan en el análisis de tendencias de datos históricos. Por ejemplo, un gerente de almacén puede utilizar pedidos anteriores del año anterior para pronosticar la cantidad de productos necesarios para los próximos 3 meses. Este enfoque permite una toma de decisiones informada y una gestión eficiente del inventario. Para modelar patrones de datos y hacer proyecciones futuras, se pueden emplear varias técnicas de series de tiempo, incluido el análisis de tendencias, la descomposición y el suavizado exponencial simple. La elección del método debe depender de si los patrones son consistentes (sin cambios en el tiempo) o en evolución (sujetos a cambios), las características de la tendencia y los elementos estacionales, y el horizonte de pronóstico deseado. Antes de generar pronósticos, es aconsejable ajustar múltiples modelos potenciales a los datos e identificar el modelo más confiable y preciso. (SOPORTE DE MINITAB 18, 2019).

El cálculo de una previsión es inevitablemente propenso a errores. Sin embargo, para mitigar estos errores, se hacen esfuerzos para seleccionar el método de pronóstico más eficaz. Por eso es crucial medir los errores. Métrica de error de pronóstico utilizada para seleccionar el mejor modelo de pronóstico.

2.2.15. Error medio (ME)

Es la diferencia entre el promedio de los valores reales y pronosticados. (Sim, 1999).

$$ME = \bar{r} - \bar{p}$$

2.2.16. Error cuadrático medio (RMSE)

Generalmente se mide la dispersión del error de pronóstico, pero esta medición amplifica el error al elevarlo al cuadrado, penalizando los períodos con mayores diferencias en comparación con otros. Como resultado, es aconsejable utilizar RMSE para períodos con desviaciones mínimas. (INGENIO EMPRESA, 2016).

$$RMSE = \sqrt{\frac{\sum \text{Error de pronóstico}^2}{n}}$$

2.2.17. Error medio absoluto (MAE)

La precisión de un pronóstico se evalúa calculando el promedio de las magnitudes de los errores de pronóstico (los valores absolutos de cada error). Cuando un analista intenta evaluar el error de pronóstico en las mismas unidades que la serie original, el método MAE resulta muy beneficioso. (sitiobigdata, 2018)

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

2.2.18. Porcentaje medio de error (MPE)

En ocasiones, resulta esencial evaluar si una técnica de pronóstico presenta sesgo, es decir, si sobreestima o subestima constantemente los resultados. En tales casos, se utiliza el porcentaje medio, que implica calcular el error para cada período, dividir ese error por el valor real para ese período específico y luego promediar estos errores porcentuales.

$$MPE = \frac{\sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i}}{n} * 100\%$$

Cuando un método de pronóstico no presenta sesgo, la ecuación MPE arroja un porcentaje que es casi cero. Por el contrario, si el resultado presenta un porcentaje significativamente negativo, indica que el enfoque de pronóstico es defectuoso. Si el resultado refleja un porcentaje positivo significativo, indica que el método de pronóstico subestima constantemente los resultados. Por el contrario, se produce una sobrestimación constante cuando el resultado es un porcentaje positivo elevado. (sitiobigdata, 2018).

2.2.19. Error absoluto medio porcentual (MAPE)

MAPE da una indicación de qué tan grande es el error de pronóstico. Comparado con el valor real de la serie. También puedes usar MAPE para comparar completamente la precisión de dos técnicas para la misma u otra técnica diferente. (sitiobigdata, 2018).

$$MPE = \frac{\sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{|y_i|}}{n} * 100\%$$

2.3. DEFINICIÓN DE TÉRMINOS BÁSICOS

2.3.1. Serie

Una serie se define como un conjunto de N observaciones o datos organizados secuencialmente y con intervalos de tiempo uniformes, que representan una característica específica a lo largo del tiempo.

2.3.2. Modelo

“Un modelo para un proceso estocástico en cualquier conjunto de hipótesis bien definidas sobre las propiedades estadísticas de dicho proceso.” (Mauricio, 2017, pág. 13)



2.3.3. Modelo Box – Jenkins

El enfoque Box-Jenkins, un método predictivo, se basa en la estimación efectiva de los parámetros mediante iteraciones del proceso.

2.3.4. Estacionariedad

La estacionariedad en una serie de tiempo implica que la función $f(t) = f(t+k)$, lo que significa que el comportamiento de la variable en el tiempo se mantiene constante incluso cuando la serie se desplaza.

2.3.5. Estacionalidad

La estacionalidad se refiere a la ocurrencia repetida de un patrón de comportamiento en intervalos regulares, como cada 3 meses, 6 meses, anualmente, cada 4 años, y así sucesivamente.

2.3.6. Modelo de pronóstico

Un modelo de predicción implica prever la ocurrencia de un evento en un período de tiempo determinado, basado en un fenómeno físico.

2.3.7. Lstm

Acrónimo de "Long Short-Term Memory", un tipo de red neuronal recurrente diseñada para modelar dependencias a largo plazo en datos secuenciales, como series temporales.



2.3.8. Red neuronal

Modelo computacional inspirado en el cerebro humano, compuesto por unidades llamadas neuronas interconectadas. Las redes neuronales se utilizan para realizar tareas de aprendizaje automático y análisis de datos complejos.

2.3.9. Entrenamiento de la red neuronal

Proceso de ajuste de los pesos y parámetros de una red neuronal utilizando datos de entrada y salida conocidos. Este proceso permite que la red aprenda a realizar tareas específicas, como la clasificación o la predicción.

2.3.10. Gestantes

Mujeres que están embarazadas en cualquier etapa de gestación.

2.3.11. Adolescente

Persona en la etapa de la adolescencia, caracterizada por cambios físicos, emocionales y sociales, generalmente entre los 13 y 19 años de edad.

2.3.12. Gestante adolescente

Adolescente que está embarazada.

2.3.13. Embarazo adolescente

Fenómeno que ocurre cuando una persona adolescente queda embarazada, usualmente entre los 13 y 19 años de edad.



2.4. OPERACIONALIZACIÓN DE VARIABLES

Tabla 1

Operacionalización de variables

VARIABLES	INDICADOR	INDICE
Variable Dependiente:	Expresada en cifras	
Numero de gestantes adolescentes en el departamento de Puno.	mensuales de gestantes adolescentes.	Gestantes adolescentes.
Variable Independiente:		
Es la misma variable dependiente comprendidos en distintos periodos de tiempo.	Periodo de registro de datos.	Meses.

Nota: Elaboración propia por el investigador.

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. POBLACIÓN

La población de estudio está conformada por el registro de gestantes adolescentes registradas en el Repositorio Único Nacional de Información en Salud (REUNIS) del Ministerio de Salud en el tiempo.

Figura 4

Mapa del departamento de Puno



Fuente: Instituto Nacional de Estadística e Informática.

3.2. MUESTRA

La selección de la muestra se realizó mediante un muestreo no probabilístico, determinado a criterio del investigador. La muestra seleccionada de gestantes adolescentes comprende desde enero de 2013 hasta diciembre de 2023 basado en los datos recopilados por el Repositorio Único Nacional de Información en Salud (REUNIS) del



Ministerio de Salud. En consecuencia, la muestra comprende estos 132 meses, abarcando el total de gestantes adolescentes registradas en el departamento de Puno durante ese período.

3.3. MÉTODO DE RECOLECCIÓN DE DATOS

Los datos necesarios para llevar a cabo esta investigación fueron obtenidos del Repositorio Único de Información en Salud REUNIS del Ministerio de Salud, abarcando información mensual desde el año 2013 hasta el 2023.

3.4. TÉCNICAS, INSTRUMENTOS, EQUIPOS Y MATERIALES

3.4.1. Técnicas de recolección

Se utilizó como técnica el fichaje, para ello se recurrió a una hoja de cálculo de Excel para registrar el número de gestantes adolescentes desde el año 2013 hasta el 2023 mensualmente.

3.4.2. Técnicas estadísticas

Se utilizó las técnicas de modelación de series de tiempo que corresponden a la metodología de Box Jenkins y una Red Neuronal Recurrente.

3.5. DISEÑO DE INVESTIGACIÓN

El diseño de investigación es no experimental.

3.6. TIPO DE INVESTIGACIÓN

El tipo de investigación para este trabajo es retrospectivo correlacional.



CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. APLICACIÓN BOX JENKINS Y UNA RED NEURONAL

Con los datos del número de gestantes adolescentes en Puno se construyeron los mejores modelos de pronóstico de tipo ARIMA y RNN, se realizaron pronósticos para los próximos meses, las evaluaciones de pronóstico con ambas metodologías fueron comparadas resultando como mejor metodología de pronóstico la Red Neuronal Recurrente.

Se exponen a continuación los datos de la serie histórica mensual correspondiente al número de gestantes adolescentes en Puno entre los años 2013 y 2023.



Tabla 2

Serie mensual de gestantes adolescentes Puno, Periodo 2013-2023

	Años										
	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Enero	14	19	30	48	26	43	27	37	20	23	16
Febrero	12	8	25	42	33	33	33	20	14	31	22
Marzo	18	11	29	50	49	54	43	32	30	40	17
Abril	17	12	22	44	44	44	37	33	26	38	24
Mayo	24	21	44	35	45	57	45	35	22	43	27
Junio	27	20	43	52	48	47	33	43	20	43	29
Julio	23	41	47	48	44	59	44	51	42	47	31
Agosto	14	43	51	53	56	54	64	49	37	48	38
Setiembre	26	43	71	60	60	58	59	50	50	59	32
Octubre	11	31	63	55	51	54	50	50	40	40	33
Noviembre	25	52	80	73	70	41	58	71	48	40	35
Diciembre	18	43	61	54	57	66	62	46	56	54	23

Nota: Elaborado por el investigador.

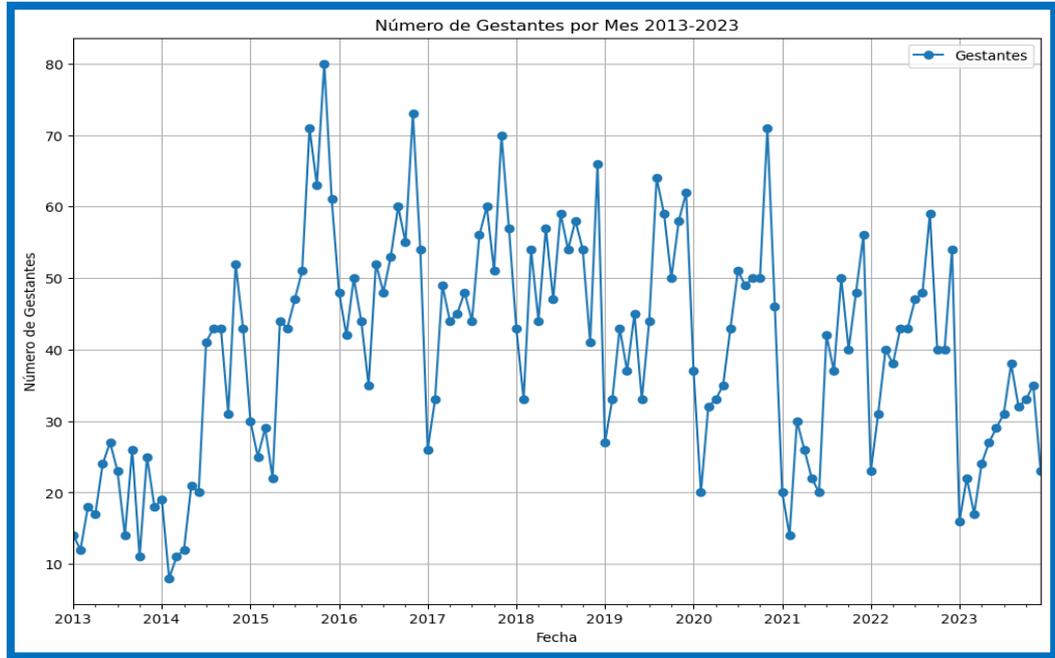
Fuente: Repositorio Único Nacional de Información en Salud (REUNIS) del Ministerio de Salud.

4.2. METODOLOGÍA BOX-JENKIS

4.2.1. Identificación del Modelo

Figura 5

Serie original de gestantes adolescente Puno periodo 2013-2023

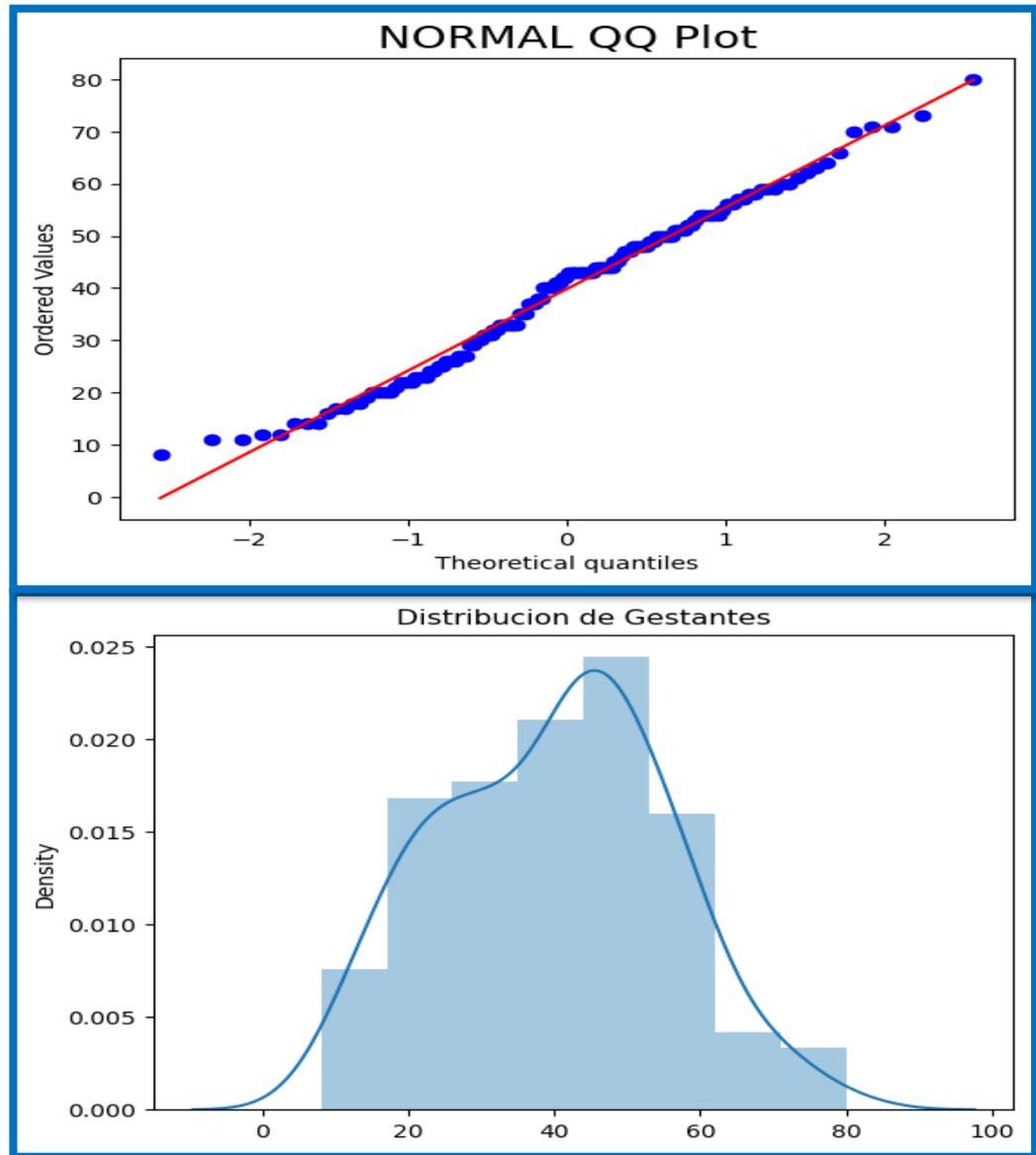


Nota: Elaboración propia con el programa Python.

En la Figura N° 5, Se puede observar los valores mensuales de la serie de gestantes adolescentes del departamento de Puno, donde se aprecia que en la serie temporal que analizamos se muestra fluctuaciones significativas desde 2013 hasta 2023, sin una tendencia clara y sin patrones obvios de repetición. Tras lo observado se evidencia que la serie no es estacionaria ni en media ni en varianza, de acuerdo al análisis gráfico.

Figura 6

Gráfico de la normalidad y distribución



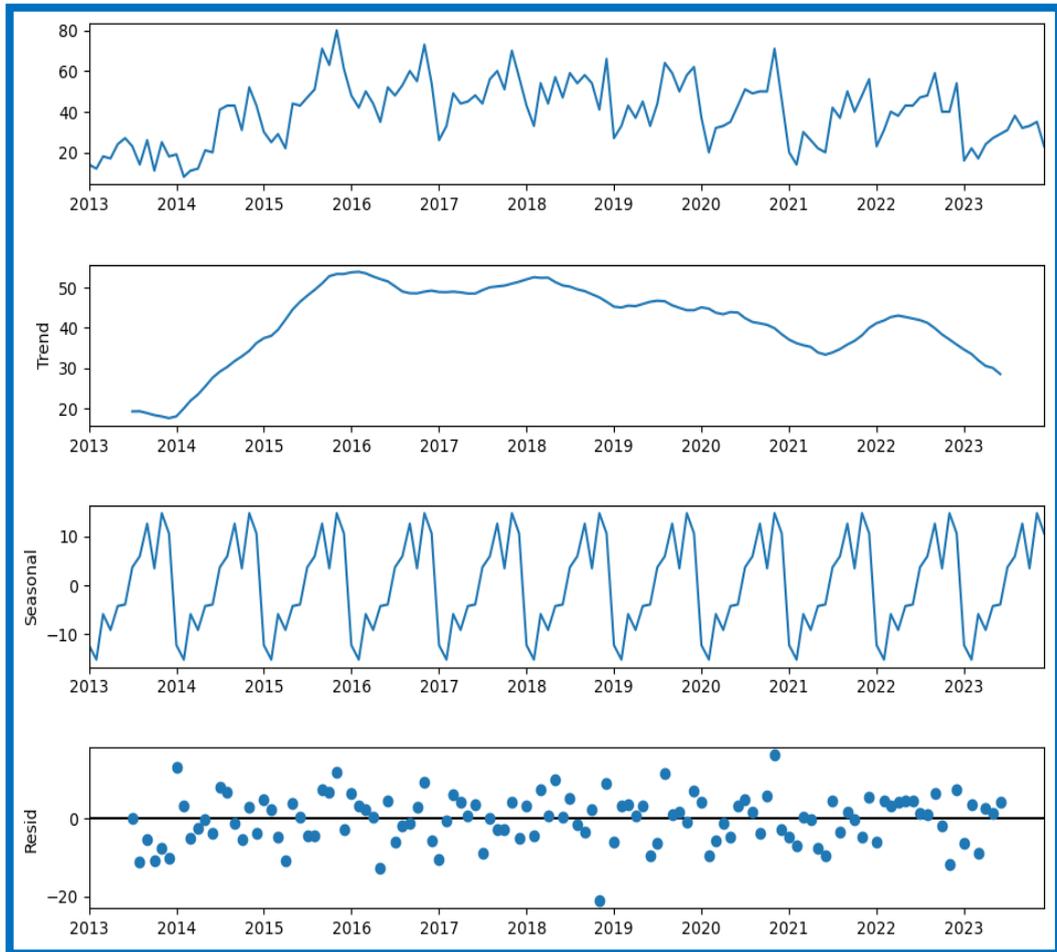
Nota: Elaboración propia con el programa Python.

En la Figura N° 6, Se puede observar que los puntos azules se desvían ligeramente de la línea roja, especialmente en los extremos. Esto sugiere que los datos no se ajustan perfectamente a una distribución normal, pero la desviación no es muy grande. En conclusión, podemos decir que los datos son aproximadamente normales, pero hay algunas desviaciones de la normalidad.

4.2.2. Gráfico de tendencia, estacionalidad y residuos

Figura 7

Gráficos de los componentes: Tendencia, estacionalidad y residuos



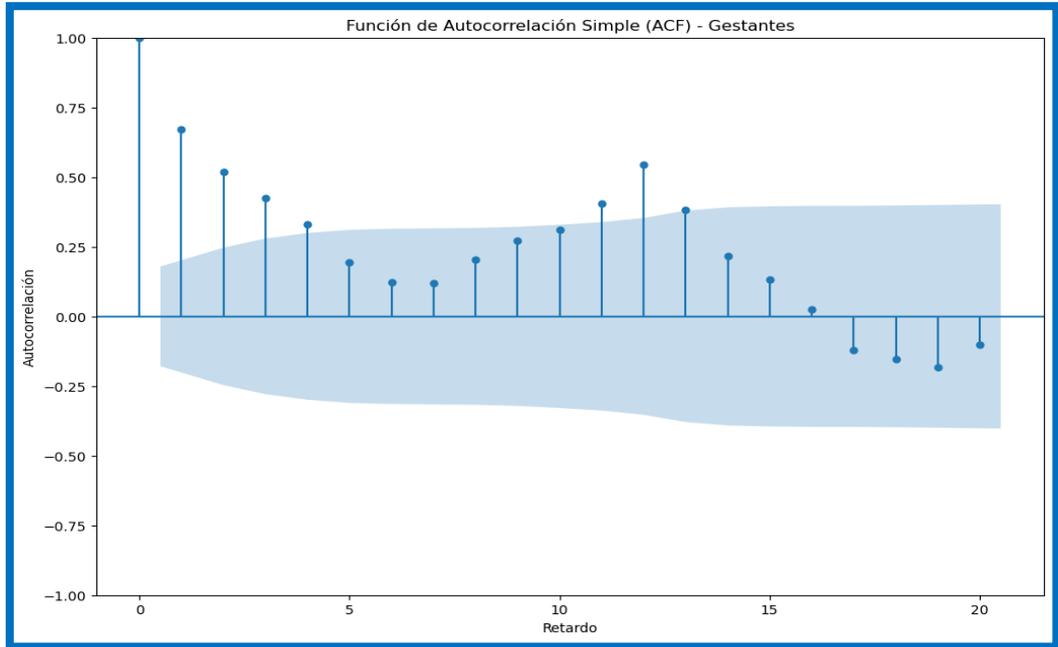
Nota. Elaboración propia con el programa Python.

En la Figura N° 7, La serie gestantes tiene una tendencia creciente hasta 2017-2018, seguida de una ligera disminución, lo que podría ser relevante para entender los factores que han influido en estos cambios. La distribución aleatoria de los residuos sugiere que el modelo de descomposición ha capturado adecuadamente la mayoría de las variaciones en la serie.

4.2.3. Función de Autocorrelación Simple y Parcial de la serie original

Figura 8

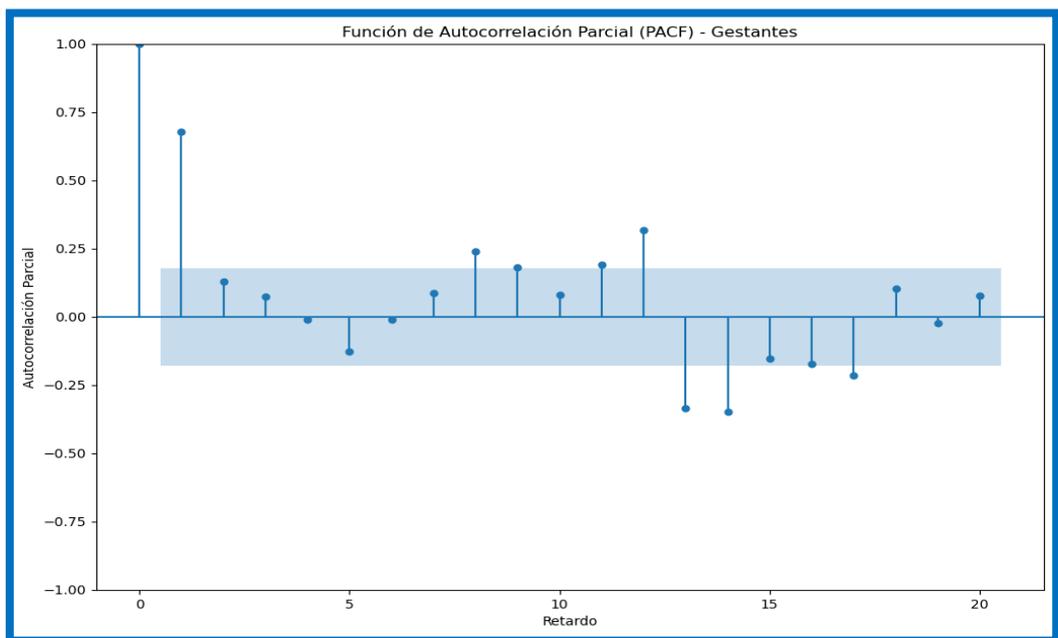
Función de Autocorrelación Simple (ACF)



Nota: Elaboración propia con el programa Python.

Figura 9

Función de Autocorrelación Parcial (PACF)



Nota: Elaboración propia con el programa Python.



En la Figura N° 8, Se muestra la Función de Autocorrelación Simple (ACF) en el cual nos muestra desfases en los retardos 1, 2, 3, 4, 11 y 12 indicando que estos se encuentran fuera de los límites de confianza.

En la Figura N° 9, Se muestra la Función de Autocorrelación Parcial (PACF) nos muestra desfases en los retardos 1, 8, 11, 13, 14, y 17 indicando que estos se encuentran fuera de los límites de confianza.

4.2.4. Prueba Dickey-Fuller Aumentada para verificar la estacionariedad

1. Hipótesis

Ho: Presenta raíz unitaria (serie no estacionaria).

Ha: No presenta raíz unitaria (serie estacionaria).

2. Regla de decisión

Si: $p. value < 0.05$, Rechazamos la hipótesis nula.

$p. value > 0.05$, No rechazamos la hipótesis nula.

Tabla 3

Prueba de Dickey Fuller A. para verificar la estacionariedad

	Valores
Test Statistic	-2.189548
p-value	0.210061
#Lags Used	13.000000
Number of Observations Used	118.000000
Critical Value (10%)	-3.487022
Critical Value (10%)	-2.886363
Critical Value (10%)	-2.580009

Nota: Elaboración propia con el programa Python.

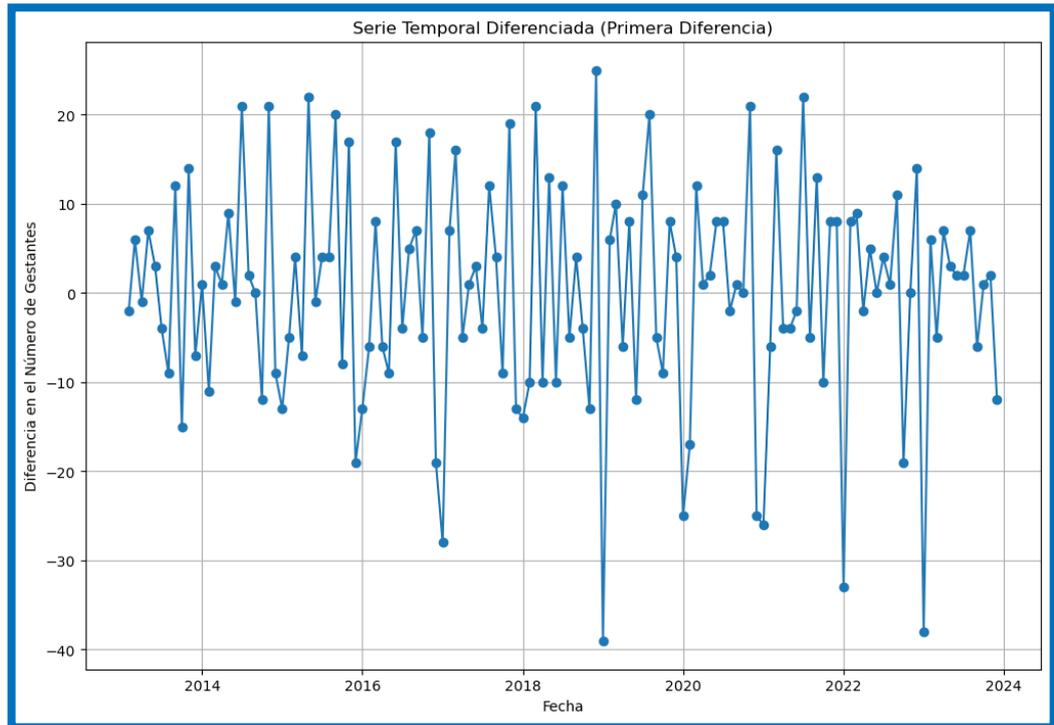
En la Tabla 3, Se observa la prueba de Dickey-Fuller Aumentada (ADF) aplicada a la serie temporal indica que no se puede rechazar la hipótesis nula de que la serie contiene una raíz unitaria, lo que sugiere que es no estacionaria. Con un p-value de 0.21, que supera el umbral estándar de 0.05, no existe suficiente evidencia estadística para declarar la serie como estacionaria. Este resultado implica que podría haber una tendencia o variabilidad en el tiempo en los datos, afectando la efectividad de los modelos que asumen estacionariedad, como los modelos ARIMA clásicos. Por tanto, se recomienda hacer la primera diferenciación de la serie, antes de proceder con el ajuste de modelos de pronóstico para mejorar la precisión y la validez de los pronósticos generados.

4.2.5. Serie diferenciada (Primera Diferencia)

Se realizó la primera diferenciación de la serie y se volverá a verificar con ADF si la serie tiene estacionariedad o estacionalidad.

Figura 10

Serie temporal diferenciada



Nota: Elaboración propia con el programa Python.

En la Figura N° 10, Se muestra una serie temporal diferenciada, muestra una variación aleatoria alrededor de cero, lo que sugiere que la serie es ahora estacionaria. Esto significa que la media, la varianza y la autocorrelación de la serie son constantes en el tiempo.

La serie temporal diferenciada también sugiere que la serie temporal original tenía un componente estacional. Esto se debe a que la serie diferenciada muestra un patrón cíclico, con picos y valles que ocurren aproximadamente en el mismo momento de cada año.



4.2.6. Prueba Dickey-Fuller Aumentada para la serie diferenciada

1. Hipótesis

H₀: Presenta raíz unitaria (serie no estacionaria).

H_a: No presenta raíz unitaria (serie estacionaria).

Alfa = 0.05

2. Regla de decisión

Si: *p. value* < 0.05 , Rechazamos la hipótesis nula.

p. value > 0.05 , No rechazamos la hipótesis nula.

Tabla 4

Dickey – Fuller A. para verificar la estacionariedad - serie diferenciada

	Valores
Test Statistic	-2.989906
p-value	0.035848
#Lags Used	12.000000
Number of Observations Used	118.000000
Critical Value (10%)	-3.487022
Critical Value (10%)	-2.886363
Critical Value (10%)	-2.580009

Nota: Elaboración propia con el programa Python.

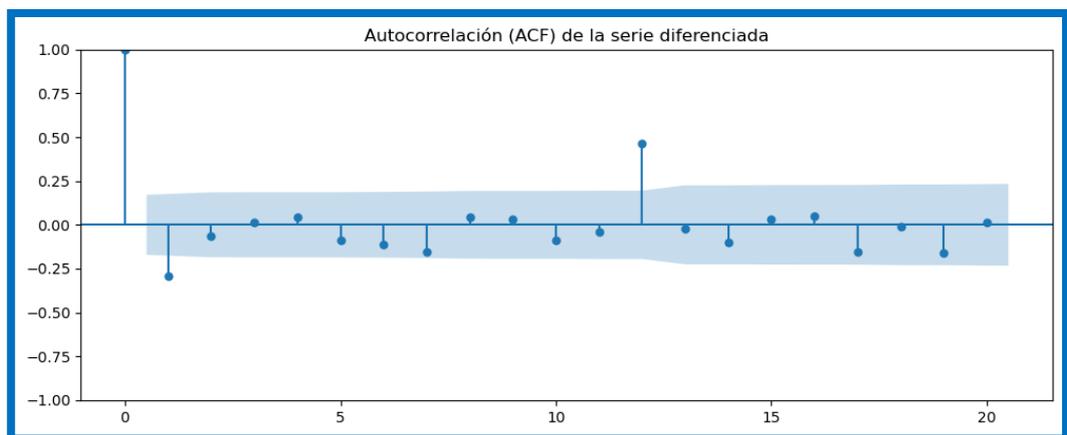
En la Tabla 4, Se observa con un valor-p de aproximadamente $0.035 < 0.05$, ahora podemos rechazar la hipótesis nula de que la serie diferenciada contiene una raíz unitaria al nivel de significancia del 5% y aceptar la hipótesis alterna. Esto indica que, después de diferenciar, la serie se ha vuelto estacionaria. El estadístico de prueba también está por debajo de los valores críticos al nivel del 5% y del 10%, reforzando esta conclusión.

4.2.7. Gráfico de ACF y PACF de la serie diferenciada

Luego de volver estacionario la serie de gestantes adolescentes en el departamento de Puno entre 2013 y 2023 se procedió a identificar los componentes de los modelos tentativos adecuados para realizar el pronóstico. Para esto, se evaluó el correlograma de la serie original diferenciada. Para identificar el modelo, es necesario evaluar las funciones de autocorrelación simple y parcial, según las figuras 10, 11. A partir de los correlogramas de la primera diferencia se determinarán modelos tentativos.

Figura 11

Autocorrelación (ACF) de la serie diferenciada

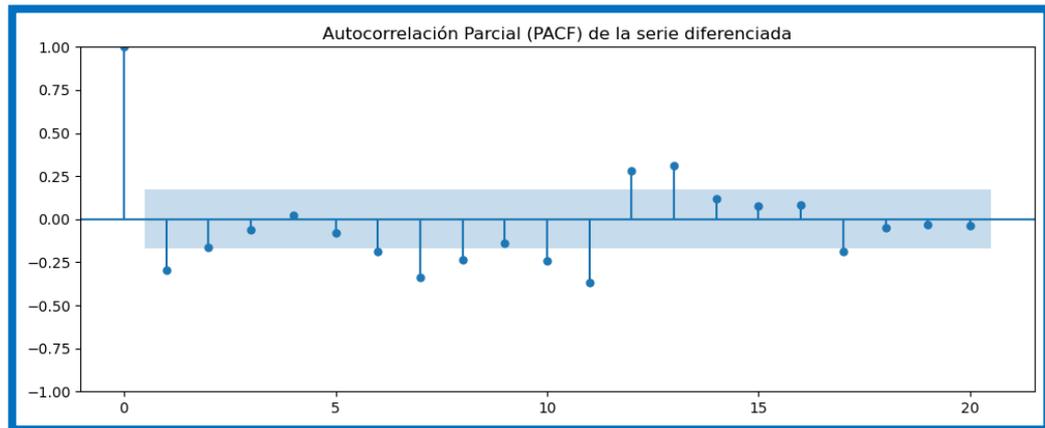


Nota: Elaboración propia con el programa Python.

En la Figura N° 11, Se muestra la Función de Autocorrelación Simple (ACF) de la serie diferenciada. La (ACF) diferenciada nos muestra desfases en los retardos 1 y 12 indicando que estos se encuentran fuera de los límites de confianza.

Figura 12

Autocorrelación parcial (PACF) de la serie diferenciada



Nota: Elaboración propia con el programa Python.

En la Figura N° 12, Se muestra el gráfico de autocorrelación parcial (PACF) de la serie diferenciada. La PACF diferenciada indica desfases en los retardos 1, 7, 8, 10, 11, 12, y 13 indicando que estos se encuentran fuera de los límites de confianza.

4.2.8. Estimación de modelos ARIMA

- ❖ Estimación de modelos ARIMA



Tabla 5

Selección de un modelo ARIMA óptimo

N°	Modelos ARIMA(p,d,q)x(P,D,Q)	Resultado (AIC)
1	ARIMA(0,1,0)(0,1,0)[12]	AIC=848.477
2	ARIMA(1,1,0)(1,1,0)[12]	AIC=810.69
3	ARIMA(0,1,1)(0,1,1)[12]	AIC=781.09
4	ARIMA(0,1,1)(0,1,0)[12]	AIC=816.46
5	ARIMA(0,1,1)(1,1,1)[12]	AIC=inf
6	ARIMA(0,1,1)(0,1,2)[12]	AIC=inf
7	ARIMA(0,1,1)(1,1,0)[12]	AIC=804.00
8	ARIMA(0,1,1)(1,1,2)[12]	AIC=inf
9	ARIMA(0,1,0)(0,1,1)[12]	AIC=811.86
10	ARIMA(1,1,1)(0,1,1)[12]	AIC=782.99
11	ARIMA(0,1,2)(0,1,1)[12]	AIC=782.99
12	ARIMA(1,1,0)(0,1,1)[12]	AIC=789.31
13	ARIMA(1,1,2)(0,1,1)[12]	AIC=784.99
14	ARIMA(0,1,1)(0,1,1)[12]	AIC=782.36

Mejor modelo: ARIMA(0,1,1)(0,1,1)[12]

Nota: Elaboración propia con el programa Python.

En la Tabla N° 5, Se han probado varios modelos ARIMA con diferentes combinaciones de los parámetros p, d, y q en el proceso de ajuste. Alcanzando un valor AIC más bajo que indicaría un modelo preferible en el cual el modelo con el menor AIC es ARIMA (0,1,1)(0,1,1)[12] con intercepto, con un AIC = 781.097 y un BIC = 789.115 lo cual sugiere que este modelo logra un buen balance entre ajuste y simplicidad.

❖ Estimación de modelos ARIMA

Tabla 6

Resultados del modelo ARIMA

Dep. Variable	y
Modelo	ARIMA(0, 1, 1) x(0,1,1,12)
Log Likelihood	-387.548
AIC	781.097
BIC	789.115
HQIC	784.347

Nota: Elaboración propia con el programa Python.

En la Tabla N° 6, Se puede observar que el modelo ARIMA(0, 1, 1) x (0,1,1,12) indica que tiene un componente AR de orden 0, una diferencia de 1 y un componente MA de orden 1 y por el otro lado se tiene un componente autoregresivo estacional AR de orden 0, una diferencia de 1 y un componente estacional MA de orden 1 y una estacionalidad periódica de 12. El modelo tiene

un valor de verosimilitud logarítmica de -387,548. El AIC del modelo es 781.097, el BIC es 789,115 y el HQIC es 784,347.

Tabla 7

Coefficientes y Significancia del Modelo

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.5856	0.071	-8.212	0.000	-0.725	-0.446
ma.S.L12	-0.8478	0.158	-5.368	0.000	-1.157	-0.538
sigma2	71.1058	10.827	6.568	0.000	49.886	92.326

Nota: Elaboración propia con el programa Python.

En la Tabla N° 7, Encontramos los coeficientes y significancia del modelo es un buen ajuste para los datos. Todos los coeficientes son estadísticamente significativos, la varianza del error del modelo es relativamente alta, lo que significa que hay una cierta cantidad de incertidumbre en las predicciones del modelo.

La siguiente ecuación del modelo ARIMA (0,1,1)(0,1,1)[12]

$$y_t = y_{t-1} - 0.5856 \epsilon_{t-1} + y_{t-12} - 0.8478 \epsilon_{t-12} + \epsilon_t$$

y_t = Valor de la serie en el tiempo t .

ϵ_t = Error en el tiempo t .

$y_{t-1,12}$ = Valor de la serie en el tiempo $t - 1, t - 12$.

$\epsilon_{t-1,12}$ = Error en el tiempo $t - 1, t - 12$.

Tabla 8

Pruebas de Diagnóstico del Modelo

	Resultados
Ljung-Box (L1) (Q)	0.04
Prob(Q)	0.84
Heteroskedasticity (H)	0.85
Prob(H) (two-sided)	0.64
Jarque-Bera (JB)	5.52
Prob(JB)	0.06
Skew	-0.23
Kurtosis	4.01

Nota: Elaboración propia con el programa Python.

En la Tabla N° 8, Según los resultados de las pruebas de diagnóstico sugieren que el modelo cumple con las condiciones de los residuos para un análisis de series de tiempo. La prueba de Ljung-Box ($Q=0.04$, $p=0.84$) no muestra evidencia de autocorrelación. La prueba de heteroscedasticidad ($p=0.64$) indica que no hay evidencia significativa de heteroscedasticidad. La prueba de Jarque-Bera ($p=0.06$) sugiere una ligera desviación de la normalidad en los residuos, con una asimetría de -0.23 . En general, estos resultados sugieren que el modelo es adecuado para el análisis de series de tiempo y que las predicciones son confiables.



4.2.9. Predicciones Box-Jenkins

Las predicciones de embarazos en adolescentes para 2024 y 2025 se muestran a continuación.

Tabla 9

Predicciones estimadas del número de gestantes adolescentes 2024

Año	Mes	Predicción
2024	Enero	7
2024	Febrero	6
2024	Marzo	15
2024	Abril	13
2024	Mayo	18
2024	Junio	18
2024	Julio	25
2024	Agosto	28
2024	Setiembre	33
2024	Octubre	26
2024	Noviembre	34
2024	Diciembre	32

Nota: Elaboración propia con el programa Python.

Tabla 10

Predicciones estimadas del número de gestantes adolescentes 2025

Año	Mes	Predicción
2025	Enero	10
2025	Febrero	9
2025	Marzo	17
2025	Abril	16
2025	Mayo	20
2025	Junio	21
2025	Julio	27
2025	Agosto	30
2025	Setiembre	35
2025	Octubre	28
2025	Noviembre	36
2025	Diciembre	33

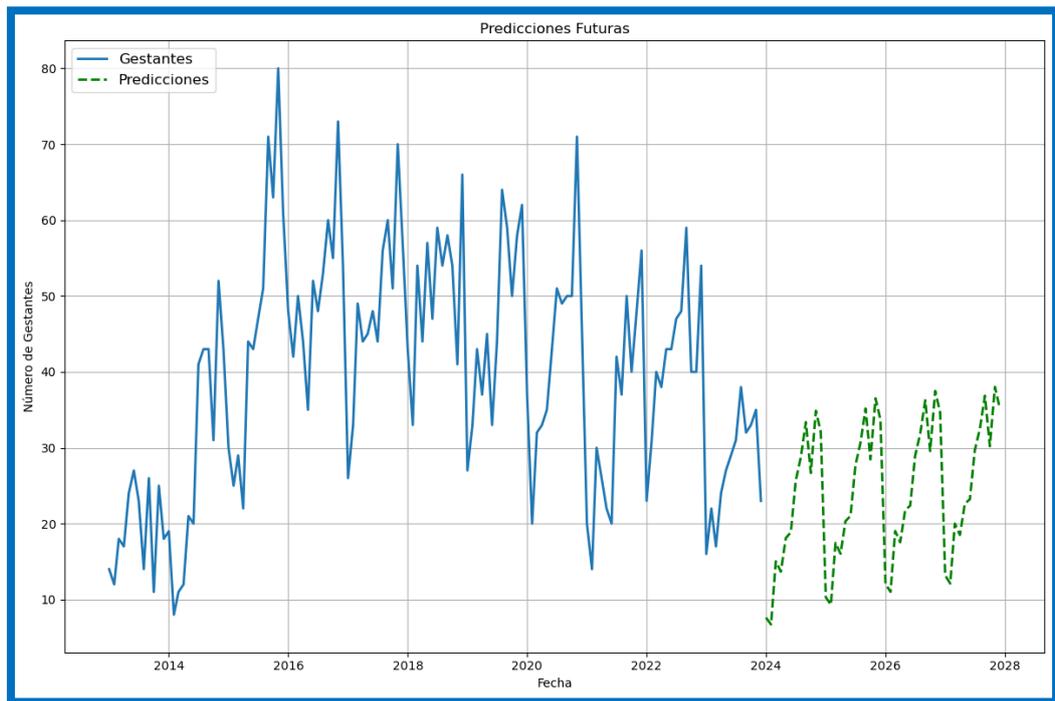
Nota: Elaboración propia con el programa Python.

En la Tabla N° 9 y 10, Los resultados predichos ofrecen una visión clara sobre la posible demanda de servicios de salud materna adolescente a lo largo del año 2024 y 2025. La identificación de picos en los meses de septiembre, noviembre y diciembre. Esta estabilidad en las predicciones puede ser útil para la

planificación de políticas y programas de salud pública destinados a abordar la gestación adolescente en esos años y posteriores, permitiendo una mejor asignación de recursos y la implementación de estrategias preventivas de manera oportuna y efectiva.

Figura 13

Predicciones futuras 2024-2025



Nota: Elaboración propia con el programa Python.

En la Figura N° 13, El gráfico muestra la validación del modelo ARIMA para predecir una serie de tiempo. La línea azul representa los datos originales de la serie, mientras que la línea verde representa los valores pronosticados. Las predicciones del modelo están relativamente cerca de los valores reales en algunas partes del período de prueba, pero hay momentos en que el modelo subestima o sobreestima significativamente los valores reales.

4.2.10. Validez del pronóstico

- Ya que tenemos el pronóstico del modelo, vamos a analizar los errores.

Tabla 11

Métrica de evaluación del Modelo

Métricas	Resultado
ME	69.59
RMSE	8.34
MAE	7.21
MPE	-25.76
MAPE	26.65

Nota: Elaboración propia con el programa Python.

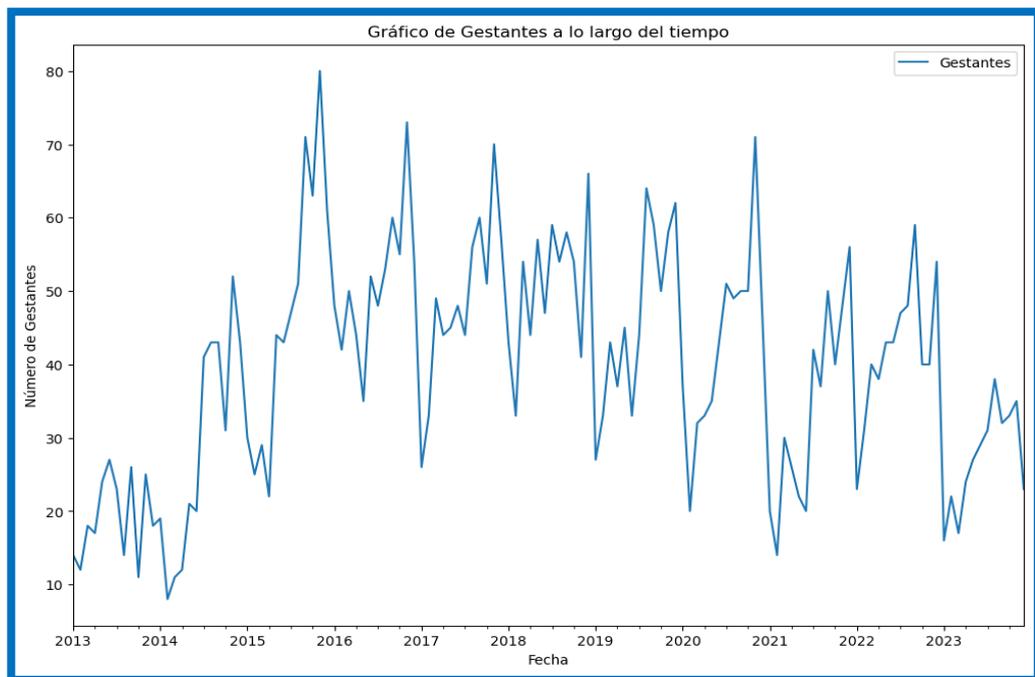
En la Tabla N° 11, El modelo de predicción muestra un error medio de 69.59, lo que indica una tendencia a sobreestimar los valores reales, aunque con un error cuadrático medio de raíz (RMSE) y un error absoluto medio (MAE) relativamente bajos de 8.34 y 7.21, respectivamente. El modelo subestima en promedio un 25.76% los valores reales (MPE), y el error porcentual absoluto medio (MAPE) es de 26.65%, lo que sugiere una precisión moderada en las predicciones, con un error promedio del 26.65% respecto a los valores reales.

4.3. METODOLOGÍA RED NEURONAL RECURRENTE (RNN)

4.3.1. Gráfico de la serie original de la serie

Figura 14

Serie original de gestantes adolescente



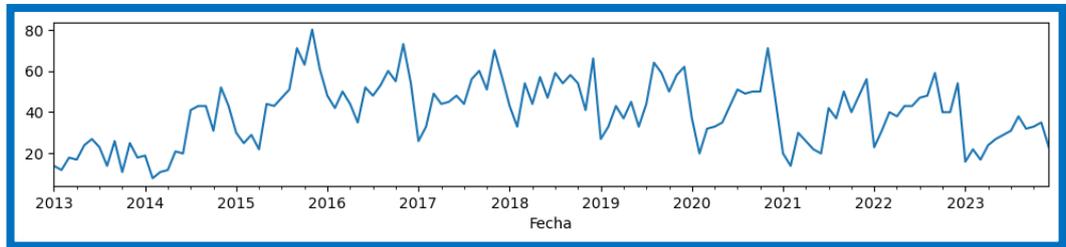
Nota: Elaboración propia con el programa Python.

En la Figura N° 14, Primero vamos a visualizar los datos originales de gestantes adolescentes en una gráfica que muestre cómo cambian con el tiempo. Luego, utilizaremos una red neuronal recurrente (RNN) para procesar estos datos y aprender de ellos de forma supervisada. Finalmente, construiremos un modelo de pronóstico óptimo utilizando esta red neuronal recurrente.

4.3.2. Descomposiciones estacionales

Figura 15

Descomposición estacional

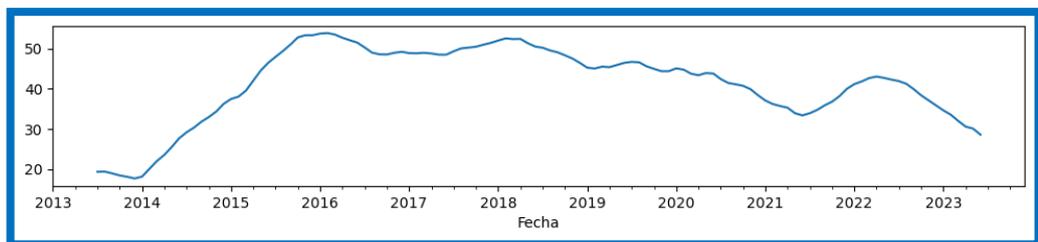


Nota: Elaboración propia con el programa Python.

En la Figura N° 15, La gráfica muestra una serie de tiempo que representa la evolución de un valor a lo largo del tiempo, desde 2013 hasta 2023. Se observa una tendencia ascendente general, con algunos picos y valles que sugieren una cierta estacionalidad o variabilidad cíclica.

Figura 16

Tendencia

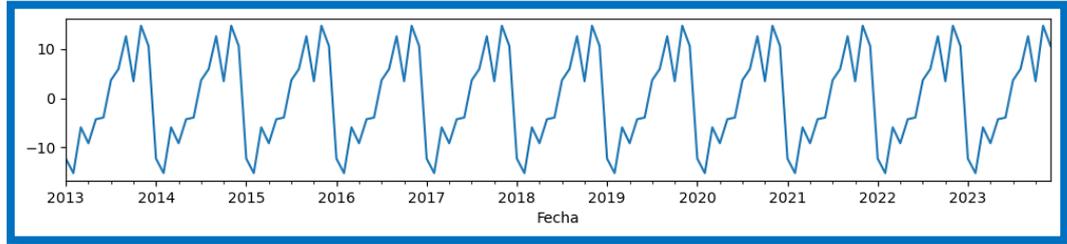


Nota: Elaboración propia con el programa Python.

En la Figura N° 16, La gráfica muestra una tendencia ascendente en los datos desde 2013 hasta 2016, luego una tendencia descendente hasta 2021, finalmente una tendencia ascendente nuevamente hasta 2022 y otra vez una tendencia descendente hasta el 2023. La serie de tiempo parece tener una estacionalidad o variabilidad cíclica.

Figura 17

Estacional

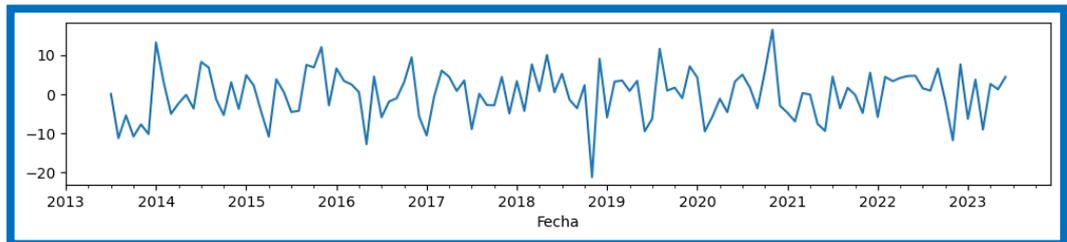


Nota: Elaboración propia con el programa Python.

En la Figura N° 17, La gráfica muestra la componente estacional de una serie de tiempo. Se observa un patrón cíclico claro, con picos y valles que se repiten aproximadamente cada año.

Figura 18

Residuos



Nota: Elaboración propia con el programa Python.

En la Figura N° 18, La gráfica muestra los residuos de un modelo de series de tiempo. Los residuos son la diferencia entre los valores reales y los valores predichos por el modelo.



4.3.3. Conjuntos de datos de entrenamiento y prueba

Tabla 12

Data de entrenamiento

Fecha	Mes	Gestantes
2013	Enero	14
2013	Febrero	12
2013	Marzo	18
2013	Abril	17
2013	Mayo	24
⋮	⋮	⋮
2022	Agosto	48
2022	Setiembre	59
2022	Octubre	40
2022	Noviembre	40
2022	Diciembre	54

Nota: Elaboración propia con el programa Python.

En la Tabla N° 12, Los datos de entrenamiento están compuesta por 120 meses desde el año 2013 - 2022, fue utilizada para entrenar el modelo de aprendizaje automático. Este conjunto permitió al modelo identificar patrones y relaciones dentro de los datos históricos, lo que le capacitó para realizar predicciones precisas sobre nuevos datos. Un conjunto de entrenamiento amplio



y representativo fue crucial para que el modelo aprendiera patrones generales y generalizara bien a datos futuros.

Tabla 13

Datos de prueba

Año	Mes	Gestantes
2023	Enero	16
2023	Febrero	22
2023	Marzo	17
2023	Abril	24
2023	Mayo	27
2023	Junio	29
2023	Julio	31
2023	Agosto	38
2023	Setiembre	32
2023	Octubre	33
2023	Noviembre	35
2023	Diciembre	23

Nota: Elaboración propia con el programa Python.

En la Tabla N° 13, Los datos de validación y prueba, que consta con los últimos 12 meses, se utilizó para evaluar el rendimiento del modelo ya entrenado.

El modelo se aplicó a estos datos, que no había visto antes, para determinar su capacidad de hacer predicciones precisas en datos nuevos. Este proceso de evaluación fue esencial para determinar la capacidad de generalización del modelo y su eficacia en escenarios reales.

4.3.4. Escalando los datos

Tabla 14

Datos de entrenamiento escalado

DATOS DE ENTRENAMIENTO ESCALADO					
0.08333333	0.48611111	0.375	..	0.70833333	0.19444444
0.05555556	0.31944444	0.61111111	..	0.58333333	0.16666667
0.13888889	0.61111111	0.55555556	..	0.69444444	0.47222222
0.125	0.48611111	0.625	..	0.75	0.40277778
0.22222222	0.30555556	0.72222222	..	0.40277778	0.58333333
⋮	⋮	⋮	⋮	⋮	⋮
0.08333333	0.48611111	0.375	..	0.70833333	0.44444444
0.05555556	0.31944444	0.61111111	..	0.58333333	0.44444444
0.13888889	0.61111111	0.55555556	..	0.37500000	0.20833333

Nota: Elaboración propia con el programa Python.

En la Tabla N° 14, Se puede observar que se hizo el escalamiento de la data entrenamiento que consiste en 120 datos para que posterior a eso se pueda generar una toma de secuencia de puntos de datos (como una serie temporal)



espaciados en intervalos iguales, teniendo en cuenta otros parámetros de series temporales, para producir batches para el entrenamiento / validación.

Tabla 15

Datos de validación escalados

N°	DATOS DE VALIDACIÓN ESCALADOS
1	[0.1111],
2	[0.1944],
3	[0.1250],
4	[0.2222],
5	[0.2638],
6	[0.2916],
7	[0.3194],
8	[0.4166],
9	[0.3333],
10	[0.3472],
11	[0.3750],
12	[0.2083])

Nota: Elaboración propia con el programa Python.

En la Tabla N° 15, Se puede observar los datos de validación/prueba escalados con un total de 12 meses para poder utilizarlo más adelante en el momento de las predicciones.

4.3.5. Creación del modelo

Tabla 16

Modelo RNN

Layer (type)	Output Shape	Param #
lstm_5 (LSTM)	(None, 100)	40800
dense_5 (Dense)	(None, 1)	101

Parámetros totales: 40901

Parámetros entrenables:

40901

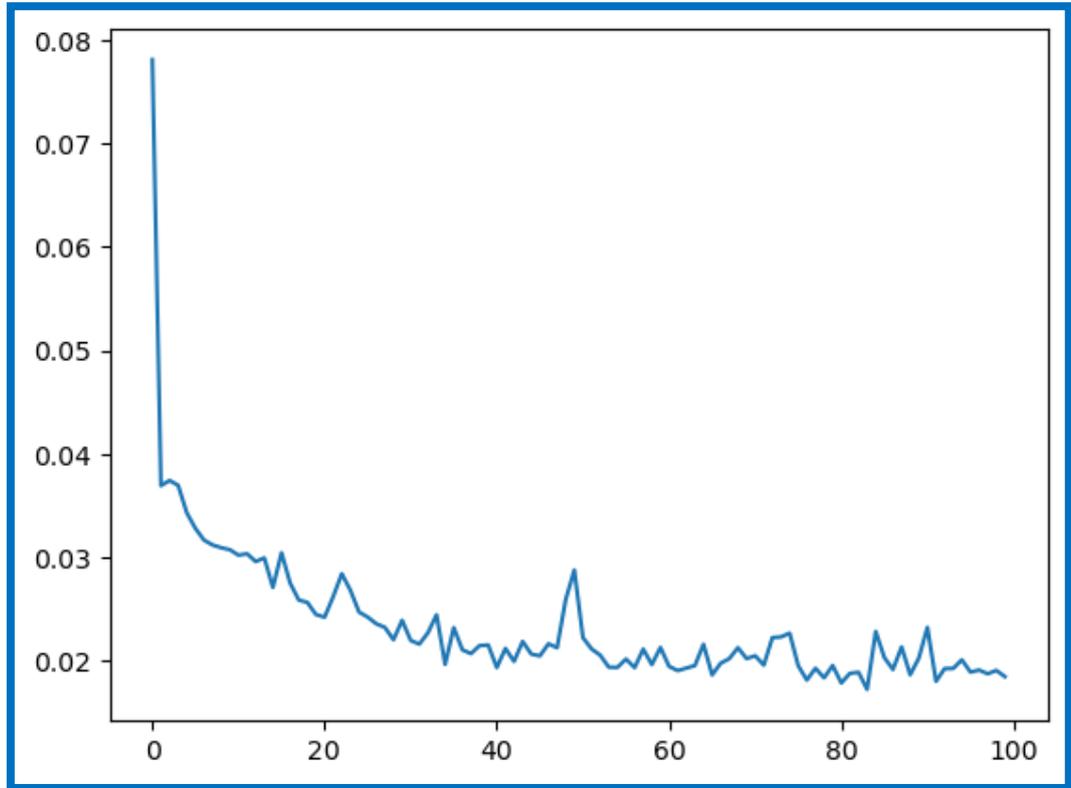
Parámetros no entrenables: 0

Nota: Elaboración propia con el programa Python.

En la Tabla N° 16, Muestra la arquitectura del modelo de red neuronal recurrente (RNN) utilizado para predecir la serie de tiempo. El modelo se compone de dos capas: una capa LSTM (Long Short-Term Memory) con 100 unidades y una capa densa con una sola unidad de salida. La capa LSTM es responsable de procesar la información temporal de la serie de tiempo, mientras que la capa densa realiza la predicción final. El modelo tiene un total de 40,901 parámetros, lo que indica una complejidad moderada.

Figura 19

Pérdida por época del Modelo



Nota: Elaboración propia con el programa Python.

En la Figura N° 19, Se muestra la pérdida del modelo durante el entrenamiento a lo largo de 100 épocas. Se puede observar que la pérdida disminuye rápidamente en las primeras épocas, lo que indica que el modelo está aprendiendo rápidamente. Después de aproximadamente 20 épocas, la pérdida se estabiliza, lo que sugiere que el modelo ha convergido y ha aprendido los patrones de los datos de entrenamiento. La pérdida se mantiene relativamente baja durante el resto del entrenamiento, lo que indica que el modelo no está sobre ajustando los datos.



4.3.6. Transformación inversa pronósticos

Tabla 17

Transformación inversa de pronósticos

N°	Transformación inversa (Pronósticos)
1	24.13609433
2	25.09040606
3	24.36138701
4	26.40550494
5	28.51991606
6	30.75058055
7	33.71258473
8	36.40692592
9	37.60010982
10	39.91805339
11	40.93401408
12	32.03672552

Nota: Elaboración propia con el programa Python.

En la Tabla N° 17, Se muestra las predicciones después de ser escalada y hacer una transformación inversa, estos datos se están preparando para hacer predicciones del modelo para que sean comparables con los valores reales de los



datos en la escala original (prueba/validación), para así poder evaluar el rendimiento del modelo.

Tabla 18

Comparación entre data test y predicción

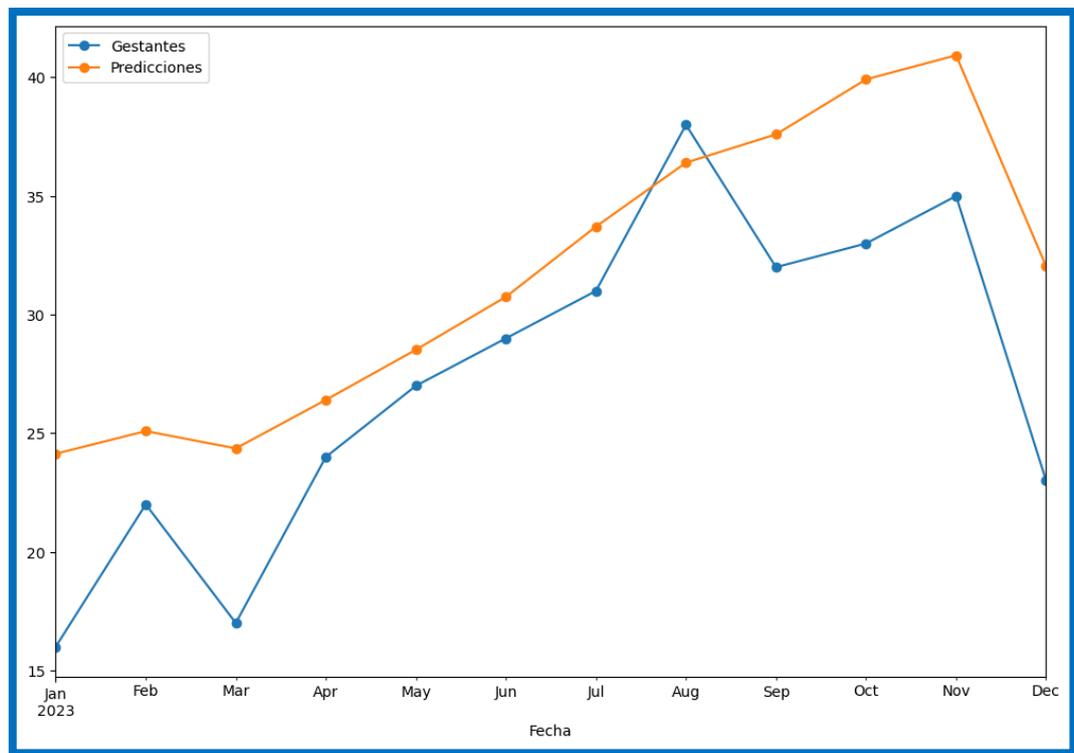
Año	Mes	Gestantes	Predicciones
2023	Enero	16	24
2023	Febrero	22	25
2023	Marzo	17	24
2023	Abril	24	26
2023	Mayo	27	28
2023	Junio	29	30
2023	Julio	31	33
2023	Agosto	38	36
2023	Setiembre	32	37
2023	Octubre	33	39
2023	Noviembre	35	40
2023	Diciembre	23	32

Nota: Elaboración propia con el programa Python.

En la Tabla N° 18, La tabla muestra las predicciones del modelo para el número de gestantes adolescentes en el año 2023, en comparación con los datos reales. Se puede observar que el modelo predice una tendencia ascendente en el número de gestantes durante el año, aunque no siempre coincide con los datos reales. Por ejemplo, en enero, marzo y diciembre, el modelo predice valores más altos que los reales. En general, el modelo parece capturar la tendencia general de los datos.

Figura 20

Comparación entre data test gestantes y predicciones



Nota: Elaboración propia con el programa Python.

La Figura N° 20, Se muestra una comparación entre los datos reales de gestantes adolescentes y las predicciones de un modelo para el año 2023. Se puede observar que el modelo captura la tendencia general de los datos, pero no es perfecto en la predicción de los valores específicos.



4.3.7. Predicciones futuras

Tabla 19

Predicciones futuras 2024

Año	Mes	Predicciones - Gestantes
2024	Enero	24
2024	Febrero	25
2024	Marzo	24
2024	Abril	26
2024	Mayo	28
2024	Junio	30
2024	Julio	33
2024	Agosto	36
2024	Setiembre	37
2024	Octubre	39
2024	Noviembre	40
2024	Diciembre	32

Nota: Elaboración propia con el programa Python.

Tabla 20

Predicciones futuras 2025

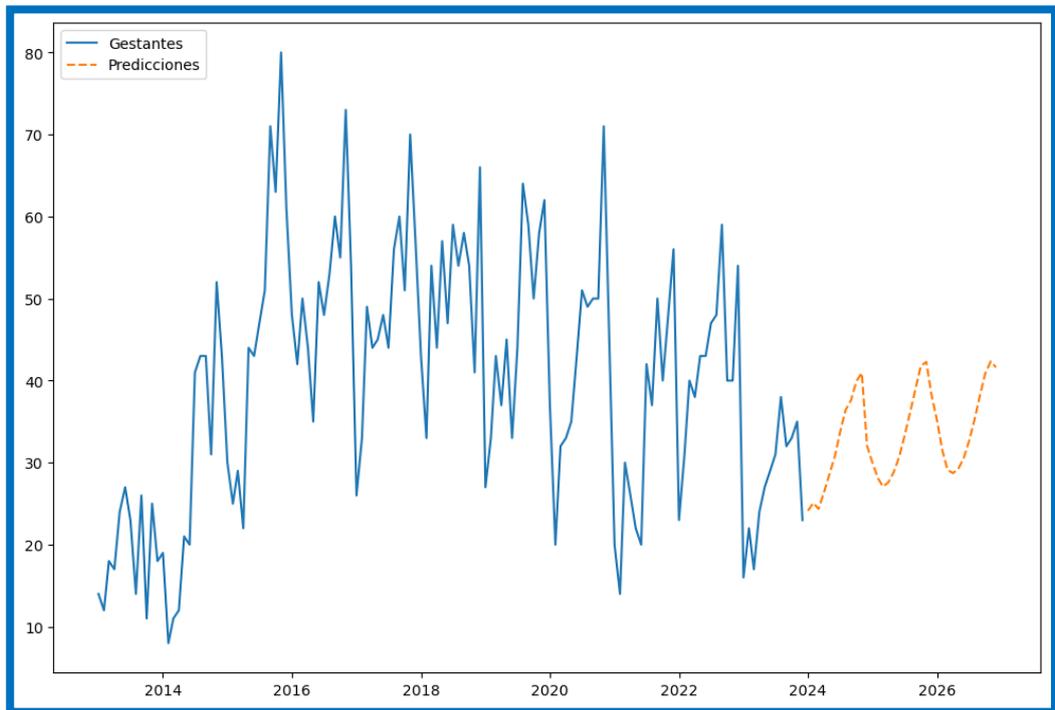
Año	Mes	Predicciones - Gestantes
2025	Enero	29
2025	Febrero	28
2025	Marzo	27
2025	Abril	27
2025	Mayo	28
2025	Junio	30
2025	Julio	33
2025	Agosto	36
2025	Setiembre	37
2025	Octubre	41
2025	Noviembre	42
2025	Diciembre	38

Nota: Elaboración propia con el programa Python.

En la Tabla N° 19 y 20, Muestra las predicciones del modelo para el número de gestantes adolescentes durante los años 2024 y 2025. Se observa una tendencia ascendente en las predicciones, lo que sugiere que el modelo espera un aumento en el número de gestantes adolescentes durante estos períodos.

Figura 21

Predicciones de los años 2024 -2025



Nota: Elaboración propia con el programa Python.

La Figura N° 21, La gráfica muestra una comparación entre los datos reales de gestantes adolescentes y las predicciones de un modelo para un período de tiempo determinado. El modelo parece ser una buena línea para hacer predicciones futuras, pero es importante tener en cuenta que la precisión de las predicciones futuras dependerá de la estabilidad de la tendencia actual y de la capacidad del modelo para adaptarse a cambios inesperados en los datos.

4.3.8. Métricas de evaluación del modelo

Tabla 21

Métricas de evaluación del modelo

Métricas	Resultados
MSE	28.92
RMSE	5.37
MAE	4.67
MPE	-19.09
MAPE	0.19

Nota: Elaboración propia con el programa Python.

En la Tabla N° 21, Las métricas de rendimiento del modelo de predicción, como el MSE (28.92), RMSE (5.37), MAE (4.67), MPE (-19.09%) y MAPE (0.19%), sugieren un buen desempeño general, con un error promedio relativamente bajo. Los valores bajos de RMSE y MAE indican una alta precisión en las predicciones, mientras que los bajos valores de MPE y MAPE sugieren que el error porcentual promedio es relativamente pequeño. En general, las métricas sugieren que el modelo de predicción tiene un buen rendimiento, con un error promedio relativamente bajo.

4.4. DISCUSIÓN

Según este trabajo de investigación, el objetivo general fue “Determinar el modelo más adecuado entre la metodología Box-Jenkins y una Red Neuronal Recurrente, para el pronóstico del número de gestantes adolescentes en el departamento de Puno durante el periodo 2013-2023”, en comparación con la investigación de (Ayala & Samillán, 2019) que su objetivo general fue “Determinar la metodología de pronóstico que ajuste mejor la serie de tiempo, precio de compra de cierre mensual de las acciones del Banco de Crédito del Perú en la Bolsa de Valores de Lima, redes neuronales artificiales o el procedimiento de Box Jenkins”.

Los hallazgos de esta investigación concuerdan con los resultados de (Ayala & Samillán, 2019) donde se encontró que las Redes Neuronales Recurrentes (RNN) superaron a la metodología Box-Jenkins para el pronóstico de series de tiempo. Ambas investigaciones concluyen que las RNN son una buena opción para pronósticos de series de tiempo, a pesar de que se aplicaron a diferentes variables (precio de acciones vs gestantes adolescentes). Sin embargo, las arquitecturas de las RNN utilizadas en ambos estudios son diferentes. El modelo RNN desarrollado en este estudio, compuesto por una capa LSTM con 100 unidades, difiere de la red neuronal artificial recurrente de tipo resiliente (1,4,1) utilizada por (Ayala & Samillán, 2019). La elección de la arquitectura puede depender de la variable que se está prediciendo y de las características específicas de la serie de tiempo.

Además, la selección del mejor modelo de Box-Jenkins también difirió entre ambos estudios. Este estudio identificó el modelo ARIMA (0,1,1)(0,1,1)[12] como el mejor modelo, mientras que Ayala utilizó un modelo SARIMA. Estas diferencias resaltan



la importancia de considerar las características específicas de la serie de tiempo al seleccionar el modelo de pronóstico más adecuado.

Estos hallazgos sugieren que se necesitan más investigaciones para explorar diferentes arquitecturas de RNN y modelos de Box-Jenkins, y para determinar la mejor estrategia para el pronóstico de series de tiempo en diferentes contextos. Futuras investigaciones deberían considerar el uso de modelos multivariantes para incluir más variables y mejorar la precisión de los pronósticos. También es importante considerar la generalización de los resultados a otros contextos y variables, así como la validación de los modelos en diferentes conjuntos de datos.

La investigación, al igual que el estudio sobre COVID-19 en Canadá de (Kumar et al., 2020) evidencia la importancia de utilizar modelos predictivos para comprender y anticipar la evolución de diferentes fenómenos. Si bien ambos estudios utilizan diferentes metodologías y variables, los resultados sugieren que las RNN, con su capacidad para capturar patrones complejos en los datos, pueden ser una herramienta valiosa para el pronóstico. (Kumar et al., 2020) utilizaron modelos de aprendizaje profundo para analizar la propagación del virus, mientras que este estudio comparó la metodología Box-Jenkins con una Red Neuronal Recurrente (RNN) para predecir el número de gestantes adolescentes. Los hallazgos de ambos estudios concuerdan en que las RNN ofrecen un mejor desempeño en términos de variabilidad de las predicciones.

Además, los resultados de esta investigación se alinean con los hallazgos de (Sagheer & Kotb, 2019) quienes propusieron un modelo DLSTM (Deep LSTM) para mejorar la capacidad de las RNN en la predicción de series de tiempo. Al igual que en este estudio, (Sagheer & Kotb, 2019) resaltan la importancia de explorar diferentes arquitecturas de modelos y de considerar la naturaleza específica de las series de tiempo



al seleccionar el modelo más adecuado. Sin embargo, es importante destacar que la elección de la arquitectura de la RNN, así como la selección del modelo de Box-Jenkins más adecuado, depende de las características específicas de la variable que se está prediciendo y del periodo de tiempo analizado.

En otra investigación de (Caira-Chuquineyra et al., 2023) indica que la prevalencia de embarazo adolescente fue del 30.9%, y se identificaron varios factores asociados como tener entre 17 y 19 años. Los factores asociados con una menor prevalencia fueron estar empleadas, estar estudiando actualmente y tener conocimiento sobre el momento del ciclo en que puede ocurrir el embarazo.

Futuras investigaciones deberían explorar la aplicación de modelos de aprendizaje profundo para predecir variables sociodemográficas, como el número de gestantes adolescentes, en diferentes contextos, y considerar el desarrollo de modelos multivariantes que incluyan más variables para mejorar la precisión de las predicciones. También es importante considerar la generalización de los resultados a otros contextos y variables, así como la validación de los modelos en diferentes conjuntos de datos.

V. CONCLUSIONES

- PRIMERA:** En base a las métricas proporcionadas, la metodología de la Red Neuronal Recurrente (RNN) resultó ser más efectiva respecto a la metodología Box Jenkins para pronosticar. La RNN presenta valores más bajos en todos los indicadores de error, lo que indica una mayor precisión para pronósticos de series de tiempo.
- SEGUNDA:** El modelo ARIMA (0,1,1)(0,1,1)[12], es el mejor modelo identificado producto del análisis de acuerdo a la metodología Box-Jenkins, presentando valores mínimos para su selección tomando en cuenta los criterios de información de Akaike (AIC) y de Bayes (BIC) obteniendo valores mínimos de $AIC = 781.097$ y $BIC = 789.115$. La ecuación del modelo es $y_t = y_{t-1} - 0.5856 \epsilon_{t-1} + y_{t-12} - 0.8478 \epsilon_{t-12} + \epsilon_t$
- TERCERA:** Se logró desarrollar un modelo de Red Neuronal Recurrente (RNN) compuesto por una capa LSTM con 100 unidades y una capa densa con una unidad de salida, con un total de 40901 parámetros entrenables. Este modelo, que aprovecha la capacidad de las capas LSTM para capturar la dependencia temporal en los datos, representa un avance significativo en la predicción.
- CUARTA:** El modelo RNN mostró un mejor desempeño en términos de variabilidad de las predicciones, con un menor error cuadrático medio (MSE) de 28.92 y un menor error cuadrático medio de raíz (RMSE) de 5.37. Además, el error absoluto medio (MAE) del modelo RNN fue de 4.67, significativamente menor que el MAE del modelo de Box-Jenkins (7.21). Sin embargo, el modelo de Box-Jenkins obtuvo una previsión razonable



en el error porcentual absoluto medio (MAPE) de 26.65, mientras que el modelo RNN obtuvo una previsión altamente precisa de 0.19, lo que sugiere una mayor precisión en la magnitud del error en relación al valor real, aunque con mayor variabilidad en las predicciones.



VI. RECOMENDACIONES

PRIMERA: Se recomienda al Repositorio Único Nacional de Información en Salud (REUNIS), que facilite el acceso a la información necesaria para llevar a cabo investigaciones periódicas. Dado que no se provee un archivo Excel u otro formato accesible que contenga los datos relevantes, se insta a que se implementen medidas que permitan a los usuarios obtener y utilizar la información de manera efectiva y constante para fines de análisis y estudio respecto a gestantes adolescentes.

SEGUNDA: Se recomienda la aplicación de otras técnicas de pronóstico y el uso de programas estadísticos como Python para el desarrollo de investigaciones en series de tiempo. Python, por su carácter de código abierto y su amplia utilización en la comunidad estadística, ofrece un entorno flexible y potente para el análisis de datos.

TERCERA: Se recomienda la realización periódica de este tipo de investigaciones para obtener datos actualizados y una visión más completa de la situación de las gestantes adolescentes.

CUARTA: Se recomienda a futuras investigaciones considerar un análisis más amplio, incluyendo variables demográficas, socioeconómicas, culturales y políticas para comprender mejor el fenómeno en estudio.



VII. REFERENCIAS BIBLIOGRÁFICAS

- Ayala, A., & Samillán, C. (2019). *Metodología de Box Jenkins vs Redes Neuronales Artificiales para construir un modelo de pronóstico del precio de compra de cierre mensual de las acciones del Banco de Crédito del Perú en la Bolsa de Valores de Lima, abril de 2005 hasta febrero de 2018*. Universidad Nacional Pedro Ruiz Gallo.
- Caira-Chuquineyra, B., Fernandez-Guzman, D., Meza-Gómez, A., Luque-Mamani, B. M., Medina-Carpio, S. L., Mamani-García, C. S., Romani-Peña, M., & Díaz-Vélez, C. (2023). Prevalence and factors associated with adolescent pregnancy among sexually active adolescent girls in Peru: Evidence from Demographic and Family Health Survey, 2015-2019. *F1000Research* 2023 11:566, 11, 566. <https://doi.org/10.12688/f1000research.108837.2>
- Calla, R. (2019). *Modelo univariante para pronosticar la cantidad de ventas mensuales de bolsas de cemento rumi producidos en la planta de cemento sur del distrito de caracoto, periodo 2005-2018*. Universidad Nacional del Altiplano.
- Coghlan, A. (2014). *A Little Book of R For Multivariate Analysis Release 0.1*. <https://people.stat.sc.edu/hansont/stat730/Coghlan2014.pdf>
- Córdova Zamora Manuel. (2006). *Estadística Inferencial*.
- Cryer, J. D., & Chan, K.-S. (2008). *Springer Texts in Statistics Time Series Analysis With Applications in R Second Edition*.
- Fournies, A. (2015). *Modelos ARMA y Box and Jenkins*. <https://doi.org/10.13140/RG.2.1.2907.0883>



- Gekko. (2017, June 24). *SARIMA (Seasonal Autoregressive Integrated Moving Average Model)*. <http://blablanegocios.com/sarima-que-es/>
- INGENIO EMPRESA. (2016). *INGENIO EMPRESA*.
<https://www.ingenioempresa.com/medicion-error-pronostico/>
- Instituto Nacional de Estadística e Informática. (n.d.). *departamento de Puno*. Retrieved October 1, 2024, from https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1205/mapas/mapa23.pdf
- Kumar, V., Chimmula, R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks R. *Chaos, Solitons and Fractals*, 135, 109864. <https://doi.org/10.1016/j.chaos.2020.109864>
- Leite, D., Monteiro, M., Monteiro, I. P., Sacramento, M., Machado, C., Vieira Bruno, Z., Aste Da Silveira, F., Francis, M., Rehme, B., Duarte Takiuti, A., Cristina, N., & Rodrigues, P. (2023). Teenage pregnancy in the first year of the COVID-19 pandemic in Brazil. *Rev Assoc Med Bras*, 69(3), 372–376. <https://doi.org/10.1590/1806-9282.20221465>
- Lewis, N. D. C. (2017). *Neural networks for time series forecasting with R : an intuitive step by step blueprint for beginners*.
- Masini, R., Medeiros, M., & Mendes, E. (2021). *Machine Learning Advances for Time Series Forecasting*.
- Mauricio, J. A. (2017). *Análisis de Series Temporales*.
<https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf>



- Melo Mayta, E. (2016). *Modelo de predicción mensual de mortalidad general intrahospitalaria en el hospital regional Manuel Nuñez Butrón-Puno, 2008-2016-I*. Universidad Nacional del Altiplano.
- Pérez Delgado, M. L., & Quintín Martín, M. (2003). *Aplicaciones de las redes neuronales artificiales a la estadística*. La Muralla.
- Pfaff, Bernhard. (2008). *Analysis of integrated and cointegrated time series with R*. 188. <https://dokumen.pub/analysis-of-integrated-and-cointegrated-time-series-with-r-2nd-ed-9780387759661-9780387759678-0387759662-0387759670.html>
- Quispe, G. (2023). *Modelo univariante de consumo mensual de energía eléctrica doméstica de la ciudad de Puno, periodo 2011 - 2021*. Universidad Nacional del Altiplano.
- Rios, G. (2008). *Series de Tiempo*.
- Rostampour, F., & Masoudi, S. (2023). Time series modeling of animal bites. *Journal of Acute Disease*, 12(3), 121. <https://doi.org/10.4103/2221-6189.379277>
- Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, 323, 203–213. <https://doi.org/10.1016/J.NEUCOM.2018.09.082>
- Sardar, I., Akbar, M. A., Leiva, V., Alsanad, A., & Mishra, P. (2023). Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: methodology, evaluation, and case study in SAARC countries. *Stochastic Environmental Research and Risk Assessment*, 37(1), 345. <https://doi.org/10.1007/S00477-022-02307-X>



Shumway, R. H., & Stoffer, D. S. (2015). *Times Series Analysis and Its Applications*.

<https://doi.org/10.1007/978-1-4419-7865-3>

sitiobigdata. (2018). *Aprendizaje Automático ml: Métricas de regresión*.

<https://sitiobigdata.com/2018/08/27/machine-learning-metricas-regresion-mse/>

SOPORTE DE MINITAB. (2019). *Forecasting with time series analysis - Minitab*.

<https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/time-series/supporting-topics/basics/forecasting/>

Villavicencio, J. (n.d.). *Introducción a Series de Tiempo*. Retrieved June 11, 2024, from

http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D

Zambrano, E., Souza, S., Quintana, S. M., Pacagnella, R. de C., & Garanhani, F. (2019).

Knowledge of Pregnant Adolescents about Human Papillomavirus. *Rev Bras Ginecol Obstet*, 41, 291–297. <https://doi.org/10.1055/s-0039-1688708>



ANEXOS

ANEXO 1 Predicciones futuras 2026

Predicciones futuras 2026

Año	Mes	RNN	BOX-JENKINS
2026	Enero	35	12
2026	Febrero	31	11
2026	Marzo	29	19
2026	Abril	28	17
2026	Mayo	29	21
2026	Junio	30	22
2026	Julio	32	28
2026	Agosto	35	31
2026	Setiembre	38	36
2026	Octubre	40	29
2026	Noviembre	42	37
2026	Diciembre	41	34

Nota: Elaboración propia con el programa Python.



ANEXO 2 Declaración jurada de autenticidad de tesis



Universidad Nacional
del Altiplano Puno



Vicerrectorado
de Investigación



Repositorio
Institucional

DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Cristian Jose Rodriguez Rojas
identificado con DNI 71451268 en mi condición de egresado de:

Escuela Profesional, **Programa de Segunda Especialidad**, **Programa de Maestría o Doctorado**
INGENIERÍA ESTADÍSTICA E INFORMÁTICA

informo que he elaborado el/la **Tesis** o **Trabajo de Investigación** denominada:
“ MODELO UNIVARIANTE BOX JENKINS Y RED NEURONAL PARA PRONOSTICAR
EL NÚMERO DE GESTANTES ADOLESCENTES, PUNO 2013 - 2022 ”

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 10 de octubre del 2024

FIRMA (obligatoria)



Huella



ANEXO 3 Autorización para el depósito de tesis en el Repositorio Institucional



Universidad Nacional
del Altiplano Puno



Vicerrectorado
de Investigación



Repositorio
Institucional

AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Cristian Jose Rodriguez Rojas
identificado con DNI 71451268 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado
INGENIERÍA ESTADÍSTICA E INFORMÁTICA

informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:

“ MODELO UNIVARIANTE BOX JENKINS Y RED NEURONAL PARA PRONOSTICAR
EL NÚMERO DE GESTANTES ADOLESCENTES, PUNO 2013 - 2022 ”

para la obtención de Grado, Título Profesional o Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los “Contenidos”) que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 10 de octubre del 2024

FIRMA (obligatoria)



Huella