



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



MODELO UNIVARIANTE DE CONSUMO MENSUAL DE
ENERGÍA ELÉCTRICA DOMÉSTICA DE LA CIUDAD DE PUNO,
PERIODO 2011 - 2021

TESIS

PRESENTADA POR:

Bach. GERBERT BEKENBAVER QUISPE QUISPE

PARA OPTAR EL TÍTULO PROFESIONAL DE:
INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2023



Reporte de similitud

NOMBRE DEL TRABAJO

MODELO UNIVARIANTE DE CONSUMO MENSUAL DE ENERGÍA ELÉCTRICA DOMÉSTICA DE LA CIUDAD DE PUNO, PERIODO

AUTOR

GERBERT BEKENBAVER QUISPE QUISPE

RECuento de palabras

19816 Words

RECuento de caracteres

105919 Characters

RECuento de páginas

120 Pages

Tamaño del archivo

1.3MB

Fecha de entrega

Aug 16, 2023 10:38 AM GMT-5

Fecha del informe

Aug 16, 2023 10:40 AM GMT-5

● 19% de similitud general

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

- 19% Base de datos de Internet
- Base de datos de Crossref
- 9% Base de datos de trabajos entregados
- 1% 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 15 palabras)

083-11804
DIR. INVESTIGACIÓN
DR. JOSE TITO LIPD.

D.N.I. 17817723

ASESOR.

DR. CONFESOR VARGAS VALVERDE

Resumen



DEDICATORIA

- *Quiero dar las gracias a Dios y a todos los miembros de mi familia dedicándoles este estudio. Mis padres son las principales personas a las que debo mis logros porque nunca me han abandonado, independientemente de las circunstancias. Les agradezco que me hayan enseñado a afrontar los retos sin dejar que me hicieran perder la compostura ni la vida.*
- *Me han inculcado las creencias, los valores, la tenacidad y la determinación que me han ayudado a convertirme en la persona que soy hoy. Todo esto lo he hecho desde el desbordamiento del amor y sin esperar recibir nada a cambio.*

Gerbert B.Q.Q.



AGRADECIMIENTOS

- *A mis docentes, sus palabras fueron sabias, sus conocimientos rigurosos y precisos, a ustedes mis profesores queridos, les debo mis conocimientos. Donde quiera que vaya, los llevaré conmigo en mí transitar profesional.*

Gerbert B.Q.Q.



ÍNDICE GENERAL

	Pág.
DEDICATORIA	
AGRADECIMIENTOS	
ÍNDICE GENERAL	
INDICE DE FIGURAS	
ÍNDICE DE TABLAS	
ÍNDICE DE ACRÓNIMOS	
RESUMEN	12
ABSTRACT.....	13
CAPÍTULO I	
INTRODUCCIÓN	
1.1. DESCRIPCIÓN DEL PROBLEMA	16
1.2. FORMULACIÓN DEL PROBLEMA	17
1.3. OBJETIVO DE LA INVESTIGACIÓN.....	17
1.3.1. Objetivo General	17
1.3.2. Objetivo Específico	17
1.4. HIPÓTESIS GENERAL	18
1.4.1. Hipótesis específicas	18
1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN.....	18
1.6. LIMITACIONES DE LA INVESTIGACION	19
CAPITULO II	
REVISION DE LITERATURA	
2.1. ANTECEDENTES DE LA INVESTIGACIÓN.....	20
2.2. BASE TEÓRICA	21



2.2.1. Técnicas de Predicción	21
2.2.2. Serie de Tiempo.....	22
2.2.3. Componentes de una Serie Temporal.....	22
2.2.4. Modelo.....	23
2.2.5. Modelo de series temporales	24
2.2.6. Ruido blanco.....	25
2.2.7. Modelo univariante.....	27
2.2.8. Componentes de una serie de tiempo	28
2.2.9. Modelos de series de tiempo	29
2.2.10. Modelo ARMA (p,q).....	30
2.2.11 Función de autocorrelación	39
2.2.12 Función de autocorrelación parcial (FACP).....	41
2.2.13. Caminata al azar	41
2.2.14. Transformación de box-cox.....	42
2.2.15. Intervalos de confianza para las predicciones	42
2.6. DEFINICION DE TERMINOS BASICOS	46
2.7. OPERACIONALIZACION DE VARIABLES.....	50
CAPITULO III	
MATERIALES Y MÉTODOS	
3.1 POBLACIÓN	51
3.2 MUESTRA	51
3.3 METODO DE RECOLECCIÓN DE DATOS	52
3.4 METODO DE TRATAMIENTOS DE DATOS	52



CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. APLICACIONES DE LA METODOLOGÍA BOX – JENKINS	62
4.1.1. Identificación del modelo	64
4.1.2. Estimación del modelo identificado	70
4.1.3. Validación o adecuación del modelo.....	72
4.1.4. Fase de predicción	76
4.3. DISCUSIÓN	78
V. CONCLUSIONES.....	82
VI. RECOMENDACIONES.....	83
VII. REFERENCIAS BIBLIOGRÁFICAS.....	84
ANEXOS.....	87

Área: Estadística

Línea: Modelos predictivos univariantes y multivariantes

FECHA DE SUSTENTACIÓN: 16 de agosto de 2023



INDICE DE FIGURAS

Figura 1: Procesos de modelos de series temporales	25
Figura 2: Series Temporales Simuladas a Partir de Varios Modelos ARIMA.....	25
Figura 3: Proceso de un Ruido Blanco	26
Figura 4: Coeficientes de autocorrelación y autocorrelación parcial de los modelos ar(1) y ar(2)	32
Figura 5: Proceso AR (1) Correlograma	33
Figura 6: Coeficientes de autocorrelación y autocorrelación parcial de los modelos MA(1) y MA (2).....	36
Figura 7: Correlograma Proceso ARMA (1,1).....	37
Figura 8: Comportamiento de la función de Autocorrelación y Autocorrelación Parcial	41
Figura 9: Metodología del enfoque BOX-JENKINS	61
Figura 10: Consumo de energía eléctrica de la ciudad de Puno, Periodo 2011 -2021 .	64
Figura 11: Función de autocorrelación estimada de la serie consumo de energía eléctrica doméstico de la Ciudad de Puno	65
Figura 12: Función de autocorrelaciones parciales estimada de la serie consumo de energía eléctrica doméstico de la Ciudad de Puno.	66
Figura 13: Función de dos diferencias no estacional por una diferencia estacional para la serie consumo de energía eléctrica doméstico de la Ciudad de Puno. ...	67
Figura 14: Función de autocorrelaciones estimadas para dos diferencias no estacional por una estacional para la serie consumo de energía eléctrica doméstico de la Ciudad de Puno.....	68



Figura 15: Función de autocorrelaciones parciales estimadas para dos diferencias no estacional y una estacional para la serie de consumo de energía eléctrica doméstico de la Ciudad de Puno.	69
Figura 16: Función de autocorrelación de residuales estimadas para la serie histórica del consumo de energía eléctrica doméstico de la Ciudad de Puno.	74
Figura 17: Función de autocorrelación parcial de residuales estimados para la serie histórica de consumo de energía eléctrica doméstico de la Ciudad de Puno.	75
Figura 18: Función de pronóstico con un 95% de confianza para la serie de consumo de energía eléctrica doméstico de la ciudad de Puno, periodo 2011 – 2021... ..	77
Figura 19: Función de pronóstico con límite de confianza 95% de consumo de energía eléctrica doméstico periodo 2011 – 2021.	78



ÍNDICE DE TABLAS

Tabla 1:	Operacionalización de variable.....	50
Tabla 2:	Series históricos del consumo de energía doméstico eléctrico	63
Tabla 3:	Resumen del modelo arima.....	71
Tabla 4:	Datos de predicción estimadas (Mw-h/mes).....	76
Tabla 5:	Valores de predicción estimadas de consumo de energía eléctrica domestico (Mwh/mes).....	88
Tabla 6:	Residuales de consumo mensual de energía eléctrica doméstico de la ciudad de Puno.....	89
Tabla 7:	Residuales estimadas de autocorrelación parcial	89



ÍNDICE DE ACRÓNIMOS

ARIMA	: Autorregresivo Integrado de Promedio Móvil.
AR	: Modelo Autorregresivo.
MA	: Media Móvil.
RSE	: Errores estándar de los residuos.
F.A.C.	: Función de Autocorrelación.



RESUMEN

El presente trabajo de investigación, comprende en la necesidad de anticiparse y proyectarse ante una demanda futura de energía eléctrica doméstico que deriva del rápido crecimiento poblacional, en el ámbito geográfico en la Ciudad de Puno de la Región Puno utilizando la metodología ARIMA (Autorregresive-Integrated-Moving Average) El objetivo es identificar modelos univariantes que permitan la descripción y la predicción datos históricos recopilados de los registros de consumo de energía existentes en electro Puno que van desde enero de 2011, hasta diciembre de 2021, para identificar el patrón, se diferenciará la serie original para que sea estacionaria, luego se usó las funciones de autocorrelación y autocorrelación parcial para identificar la forma del modelo, estimado, para validar el modelo, se realizó el análisis de los residuos, con lo que se verifico que los residuos sean compatibles con el ruido blanco, utilizando la prueba aumentada de Dickey Fuller para predecir el mejor conjunto univariante integrado. En conclusión, utilizando la metodología Box-Jenkins para lograr los objetivos de la investigación, el conjunto nos proporcionará mejores tendencias de crecimiento que no muestren signos de cambios cíclicos y estacionales, y finalmente el mejor modelo unidimensional que nos permita describir y predecir el comportamiento del consumo eléctrico doméstico en el modelo univariado del conjunto.

Modelo Multiplicativo

ARIMA (2,1,1) (0,1,1)₁₂

$$\hat{Y}_t = 2Y_{t-1} - Y_{t-2} + Y_{t-12} - Y_{t-13} - (0.851407)\varepsilon_{t-1} - (0.904686)\varepsilon_{t-12} + 0.770255\varepsilon_{t-13}$$

Palabras Clave: Modelo Univariante, Consumo, Energía Eléctrica, Ciudad de Puno.



ABSTRACT

This research work includes the need to anticipate and project a future demand for domestic electricity that derives from rapid population growth, in the geographical area in the City of Puno in the Puno Region using the ARIMA (Autoregressive-Integrated - Moving Average) The objective is to identify univariate models that allow the description and prediction of historical data collected from existing energy consumption records in electro Puno that go from January 2011 to December 2021, to identify the pattern, it will be differentiated the original series to be stationary, then the autocorrelation and partial autocorrelation functions were used to identify the shape of the model, estimated, to validate the model, the analysis of the residuals was carried out, with which it was verified that the residuals are compatible with white noise, using the augmented Dickey Fuller test to predict the best integrated univariate ensemble. In conclusion, using the Box-Jenkins methodology to achieve the research objectives, the set improved us the growth trends that do not show signs of cyclical and seasonal changes, and finally the best one-dimensional model that allowed us to describe and predict the behavior of the domestic electricity consumption in the univariate model of the ensemble.

Multiplicative Model

ARIMA (2, 1, 1) (0, 1, 1)₁₂

$$\hat{Y}_t = 2Y_{t-1} - Y_{t-2} + Y_{t-12} - Y_{t-13} - (0.851407)\varepsilon_{t-1} - (0.904686)\varepsilon_{t-12} + 0.770255\varepsilon_{t-13}$$

Keywords: Univariate Model, Consumption, Electric Power, City of Puno.



CAPÍTULO I

INTRODUCCIÓN

Una actividad económica importante a escala nacional es la generación de electricidad doméstica, y se explorará la necesidad de pronosticar y predecir el rápido aumento de la demanda de electricidad en el futuro, generando así más electricidad para el consumo de electricidad. crea la necesidad de predecir el suministro eléctrico adecuado, para lo cual se propone establecer un modelo univariado para describir y predecir el consumo eléctrico mensual.

Los datos fueron recopilados de los registros existentes proporcionado por el servicio de Electro Puno, División de Operaciones Comerciales, Electro Puno S.A.A. Por un periodo desde 2011 – 2021, Al igual que ocurre con la agrupación mensual, Para llegar a una conclusión sobre la variable objeto de esta investigación, es necesario realizar un análisis de la conducta de la serie histórica. La utilización de los valores para crear un tipo de modelo que capte suficientemente la conducta de la cambiante en cuestión en el pretérito y realice presagios adecuadas utilizando técnicas ARIMA estocásticas necesarios para el análisis de series temporales univariantes. Resultó ser una de las técnicas más precisas para predecir cuantitativamente el futuro.

Cuando se tomó la decisión el investigador, a menudo te encuentras en un ambiente de incertidumbre sobre posibles eventos futuros. En cualquier caso, si el investigador logra reducir en cierta medida la incertidumbre de los eventos futuros, puede lograr mejores resultados. Para reducir la incertidumbre futura, se identificaron modelos univariados de predicción del futuro o predicciones basadas en datos históricos, que luego se procesaron mediante el método de Box-Jenkins.



La tesis consta de cinco capítulos:

Capítulo 1 Consideraciones de carácter general en este capítulo se abordan la normalización de la cuestión, los motivos del análisis, la prueba de los análisis, las restricciones de la observación, las hipótesis y las variables, se expone la razón de ser de la tesis.

Capítulo 2 Es el marco teórico, que considera los antecedentes del estudio, los fundamentos teóricos y las definiciones de los términos de este capítulo, además explica las referencias a los temas tipo tesis.

Capítulo 3 Cubre la metodología, el tipo de estudio, entorno del estudio, población y muestra, métodos y herramientas de recopilación de datos, validez y confiabilidad de las herramientas, diseño y detalles de la recopilación y el procesamiento de datos.

Capítulo 4 Resultados, presenta los resultados del procesamiento de datos y describe la discusión de fondo.

Capítulo 5 Conclusiones, describe todas las conclusiones del estudio tal como se describen en los objetivos.



1.1. DESCRIPCIÓN DEL PROBLEMA

En general, se considera que los problemas incluyen la necesidad de anticiparse y proyectarse la futura demanda de electricidad de la población en rápido crecimiento de Puno, lo que resulta en un mayor consumo doméstico de electricidad, lo que da como resultado una demanda pronosticada para proporcionar suficiente energía para toda la ciudad de Puno, que es la base para completar este trabajo de investigación.

El acceso a servicios básicos antes inexistentes, como la electricidad doméstica, es fundamental, convirtiéndose en una demanda latente que ya no se limita a los hogares urbanos porque el suministro eléctrico doméstico es suficiente. Para lograr este objetivo, se requieren servicios básicos de alta calidad y acordes a las necesidades de la población, lo cual es determinante para su desarrollo. Ello se debe principalmente a que el acceso a servicios básicos antes inexistentes, como la electricidad doméstica, es fundamental. En los últimos años, la población ha aumentado significativamente, lo que se traduce en un alto consumo de energía eléctrica en los hogares, lo que lleva a predecir el futuro, lo que ha llevado a la empresa comercializadora Electro Puno a brindar posibles eventos futuros durante muchos años, pero luego lograr mucho. resultados confiables. Captando las actividades de la ciudad de Puno, las diferencias en el comportamiento de los consumidores domésticos de energía eléctrica explican el desarrollo socioeconómico, tratando de obtener un modelo predictivo que defina el promedio general de eventos futuros. Por otro lado, carece de información relacionada con la previsión del consumo eléctrico.

El sistema eléctrico debe brindar un servicio de calidad a todos los usuarios, por lo tanto, al momento de planificar el sistema en detalle, debe funcionar un sistema eléctrico confiable, que permita a los usuarios conocer el estado actual y qué hacer en nuevas condiciones futuras. Una herramienta útil en la planificación del sistema



energético es la previsión del consumo de energía, que proporciona una vista previa de la expansión del sistema energético; para predecir las mejoras del servicio se convierte en el primer paso en algún desarrollo de planificación del método eléctrico.

1.2. FORMULACIÓN DEL PROBLEMA

Frente a esta problemática de vital importancia sobre el consumo de energía eléctrica frecuentes con miras a contribuir con el conocimiento para el análisis y proyección de datos futuros, en la búsqueda de una solución inteligente al planteamiento del problema y dar alternativas de prevención a los problemas prioritarios del servicio eléctrico de la región. Con estas cuestiones en mente, se pueden hacer la siguiente pregunta:

¿Cuál es el modelo univariado que permita describir y predecir los patrones de consumo mensual de energía eléctrica doméstico en Puno - Electro Puno para el período 2011 al 2021?

1.3. OBJETIVO DE LA INVESTIGACIÓN

1.3.1. Objetivo General

Determinar el ajuste de un modelo univariado de series de tiempo que permita describir y pronosticar el consumo energético doméstico en la ciudad de Puno - Electro Puno del 2011 al 2021.

1.3.2. Objetivo Específico

- Describir, el comportamiento del consumo eléctrico domiciliario de la ciudad de Puno - Electro Puno, periodo 2011-2021.
- Validar el modelo estimado para el consumo de energía eléctrica doméstico de la ciudad de Puno Electro Puno, periodo 2011 – 2021.



- Pronosticar con el mejor modelo estimado, el consumo eléctrico doméstico en la ciudad de Puno - Electro Puno para el período 2011-2021.

1.4. HIPÓTESIS GENERAL

El modelo univariado ARIMA multiplicativo de Box-Jenkins proporciona un mejor modelo de ajuste que un modelo no integrado de Box-Jenkins en el consumo mensual de energía eléctrica doméstico de la ciudad de Puno para el período 2011-2021.

1.4.1. Hipótesis específicas

- La serie de consumo de energía eléctrica en Puno mostró una tendencia creciente y positiva efectiva sin cambios cíclicos y estacionales.
- Los modelos univariados integrado multiplicativo dan un mejor ajuste y pronóstico, se adaptan mejor a la serie de consumo eléctrico doméstico 2011 - 2021 de la ciudad de Puno - Electro Puno.

1.5. JUSTIFICACIÓN DE LA INVESTIGACIÓN

Las previsiones del gasto energético reflejan los requerimientos venideros de la comunidad. Esta anticipación debe reflejar mejor la realidad, ya que los valores inferiores a los reales crean lagunas en la presentación de los servicios futuros, y los valores superiores a los reales fomentan la inversión en instalaciones que no se utilizan inmediatamente.

En los últimos años, la demanda de electricidad ha ido en aumento. La mejora del nivel de vida en Puno está asociada a una mayor demanda de energía eléctrica domiciliaria, ya que se utilizan más los dispositivos electrónicos y el acceso a estos servicios se ha convertido en una necesidad, al igual que el agua.



Con el fin de definir el comportamiento del consumo eléctrico y realizar predicciones futuras utilizando métodos estadísticos alternativos adecuados que permitan el análisis de la información obtenida, se decidió ajustar los datos al consumo eléctrico de los hogares mediante la formulación de un modelo en función del tiempo. análisis de series para inferir de Describir en un modelo, luego explicar y tomar decisiones apropiadas.

Para lograr un mejor servicio, se espera encontrar un modelo de pronóstico mensual que se adapte al consumo eléctrico de los hogares, para que los habitantes puedan recibir un servicio de calidad y beneficios, siendo la electricidad utilizada en los hogares el elemento básico de producción. En varios sectores incluyendo industrial, comercial y residencial o privado. Los beneficios de mejores servicios (principalmente electricidad domiciliaria) contribuirán así al desarrollo social y comercial de la población, y será de interés colectivo planificar activamente mejores servicios para la población.

1.6. LIMITACIONES DE LA INVESTIGACION

Una de las limitaciones del proyecto de investigación fue la dificultad de recolectar datos históricos de varios meses debido a que se encontraba en diferentes libros y no había datos completos, por lo que solo se obtuvo información que pudiera ser utilizada para el análisis y evaluación. proyectos de desarrollo de la investigación.



CAPITULO II

REVISION DE LITERATURA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

Orbezo, H. A. (2011) aplicó: El análisis de series de tiempo según el método estocástico ARIMA (modelo estocástico lineal integrado autorregresivo – promedio móvil) para el pronóstico de la demanda eléctrica residencial (con tendencia, ciclos estacionales y estocasticidad) en la región sur de Lima ha demostrado ser efectivo para aumentar la eficiencia y el pronóstico exactitud. comparar y validar los resultados obtenidos utilizando métodos de predicción deterministas. La determinación del análisis estocástico ARIMA que se utilizará para la modelación y anticipación de la petición de energía de la región poblacional de Sur Lima. El MODELO ARIMA $(p,d,q) \times (p,d,q)$ de consumo de energía incluye la constante.

Juculaca, (2019) Tesis título: “Modelo univariante para predecir el número de casos de infecciones respiratorias agudas, neumonía y defunciones en niños menores de 5 años en la dirección regional de salud Puno – 2018”. Estableció que un excelente ejemplar de predicción identificados son: SARIMA $(2,0,0)(0,1,1)_{12}$, SARIMA $(1,0,1)(2,0,0)_{12}$ Y SARIMA $(2,0,1)(2,0,1)_{12}$, que expresan las secuencias invariable estacionales.

Carcasi (2017) encontró: El modelo univariante para el consumo mensual de energía eléctrica doméstica en el distrito de Putina – Electro Puno, periodo 2005 – 2015. Es ampliamente reconocido que el problema es la necesidad de predecir y predecir la demanda futura de electricidad en rápido crecimiento, lo que dará como resultado un aumento en el consumo de electricidad, y, por lo tanto, el objetivo es: identificar modelos unidimensionales que permitan la descripción y predicción. de energía eléctrica mensual.



consumo de energía. Los datos se recopilaron de los registros de consumo de electricidad existentes. El modelo implementado para describir y ajustar los datos es un modelo ARIMA multiplicativo. ARIMA (0,2,1)(0,1,1).

Leonardo (2017) halló: El Modelo univariante para el consumo doméstico mensual de agua potable en el distrito de Ilave – EMSA Puno, periodo 2002 – 2013. Se ha determinado que el modelo que proporciona la mayor coincidencia es un modelo unidimensional que permite anticipar y prever el comportamiento del agua potable extraída del grifo mensualmente utilizando los datos de EMSA Puno. Funciona de acuerdo al siguiente modelo Modelo ARIMA (1,1,1)(0,1,1).

Arramatia (2010) concluyó que: Los modelos univariantes que mejor se ajustan para decidir y predecir el comportamiento de la serie de la Energía Eléctrica (kw/mes), en la ciudad de Juliaca, es ARIMA (1,1,1), y el número de usuarios de Energía Eléctrica periodo 2004 – 2009, es ARIMA (0,2,1): $\Delta^2 y_t = (1 - \theta_1 \beta) a_t$

2.2. BASE TEÓRICA

2.2.1. Técnicas de Predicción

El pronóstico es el uso de datos pasados sobre una variable para predecir su desempeño futuro. La primera proyección realizada se basa en los mismos datos utilizados para realizar y calibrar el modelo de muestra, mientras que la segunda proyección se basa en datos externos de la muestra. Los pronósticos estáticos son pronósticos basados en la última información disponible. El pronóstico dinámico se caracteriza por el uso de los últimos pronósticos disponibles como datos para el próximo pronóstico, lo que permite realizar pronósticos para dos o más períodos futuros.

2.2.2. Serie de Tiempo

Una serie temporal se basa en información sobre la evolución de las variables a lo largo del periodo. En medio de ellos, el modelo de Box-Jenkins conforma un grupo de procedimientos para un procesamiento de secuencias temporales. Es un grupo de observaciones que se organizan de acuerdo con las características cuantitativas de los fenómenos individuales. En diferentes momentos, las observaciones en la serie de tiempo deben estar en orden de mayor a menor porque la mayor parte de la información se perderá, y dado que estamos interesados en descubrir cómo cambian las variables con el tiempo, el orden cronológico es importante. de las observaciones.

2.2.3. Componentes de una Serie Temporal

Córdova (2006) vienen hacer las siguientes:

Tendencia

Es una parte vital de la línea de tiempo, ya que indica su desarrollo a lo largo del tiempo. Se muestra como una línea recta perpendicular al eje x y puede permanecer inmóvil o prácticamente constante.

Funciones Cíclicas

A lo largo de la línea que indica la tendencia general, podemos ver varios movimientos alcistas y negativos a corto plazo. Se les ha dado este apodo debido a que sus secuencias se repiten como ruedas que giran. El símbolo C representa el componente de volatilidad de la expresión.

Variaciones Estacionales

Es el nombre que se le da a las fluctuaciones anuales que tienen aproximadamente la misma forma de un año a otro. Dependiendo de la naturaleza



de la serie, los períodos oscilantes pueden ser tan cortos como horas, días, semanas, meses o trimestres, pero no pueden exceder de un año. El componente de variación estacional se denota por E.

Movimientos Irregulares

Se trata de cambios de tendencia por causas aleatorias o esporádicas (como huelgas, inundaciones, etc.), no puede atribuirse a efectos estacionales o cíclicos. Los componentes que cambian irregularmente están marcados con I.

2.2.4. Modelo

La representación formal de una teoría o la representación matemática de los hechos que se han visto es lo que se conoce como modelo.

Los resultados del análisis estadístico del modelo se presentan en un formato matemático. Para construir un modelo hay que tener acceso a una gran base de datos de observaciones, así como experiencia, intuición, creatividad, sencillez y capacidad para elegir el subconjunto de variables con el menor número posible de grados de libertad.

Modelos deterministas: estos son métodos simples de extra población en los que la secuencia de referencia no tiene una fuente subyacente o una naturaleza de aleatoriedad, y su relativa simplicidad suele ir acompañada de una menor precisión.

Modelos estocásticos: basados en una descripción simplificada del proceso estocástico subyacente en una secuencia; en términos simples, tome la secuencia observada: Y_1, Y_2, \dots, Y_t ; se deriva de un conjunto de variables aleatorias difíciles de determinar con alguna distribución común, creando así un modelo aproximado útil para generar predicciones.

2.2.5. Modelo de series temporales

Al realizar análisis de series temporales, se aplican como herramientas analíticas teorías deterministas o estocásticas, o una mezcla de ambas.

Las variables que se miden en distintos momentos del tiempo se denominan variables temporales. El valor de la variable "Y" en el momento indicado por "t" se denota por la variable "Y_t".

Serie temporal: Se refiere a la colección de "t" observaciones, con un valor asignado a cada una de las variables Y₁, Y₂, ..., Y_t. Otro nombre para esto es serie temporal.

Existen tres modelos diferentes para las series temporales que suelen considerarse aproximaciones adecuadas de las conexiones reales que se han observado entre los distintos componentes de los datos.

Aditivo:
$$Y(t) = T(t) + E(t) + C(t) + A(t)$$

Multiplicativo:
$$Y(t) = T(t) \times E(t) \times C(t) \times A(t)$$

Mixto:
$$Y(t) = T(t) \times E(t) - C(t) + A(t)$$

Donde:

$Y(t)$: Serie observada en Instante t .

$T(t)$: Componente de Tendencia.

$E(t)$: Componente Estacional.

$C(t)$: Variaciones Cíclicas.

$A(t)$: Componente Aleatoria (accidental).

Una suposición común es que $A(t)$ es un componente aleatorio o ruido blanco con media cero y varianza constante.

Figura 1:
Procesos de modelos de series temporales

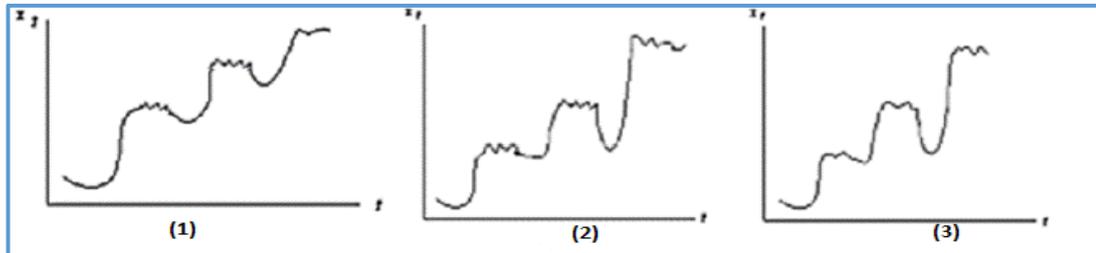
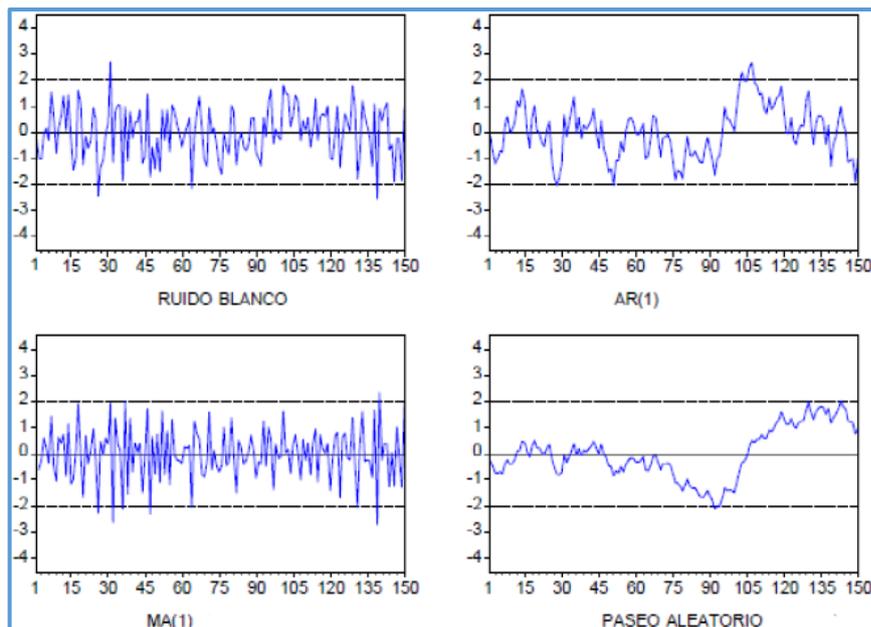


Figura 2:
Series Temporales Simuladas a Partir de Varios Modelos ARIMA



2.2.6. Ruido blanco

Es un signo estocástico (progreso aleatorio) caracterizada por el hecho de que sus valores de señal en dos momentos diferentes no están estadísticamente correlacionados en el tiempo. Por ello, su densidad espectral de potencia, también conocida como PSD, es siempre la misma, y su gráfica es siempre la misma, lo

que indica que la señal comprende todas las frecuencias y que todas las frecuencias tienen el mismo impacto que la luz blanca.

El ruido blanco es un proceso aleatorio más simple llamado ruido blanco, que es una secuencia de variables aleatorias con media cero, varianza constante y covarianza cero, se denotará habitualmente por $a_t; t = 0; \pm 1, \pm 2, \dots \infty$; una variable a_t se denomina Ruido Blanco si se cumple las siguientes condiciones (González,2009,p.18).

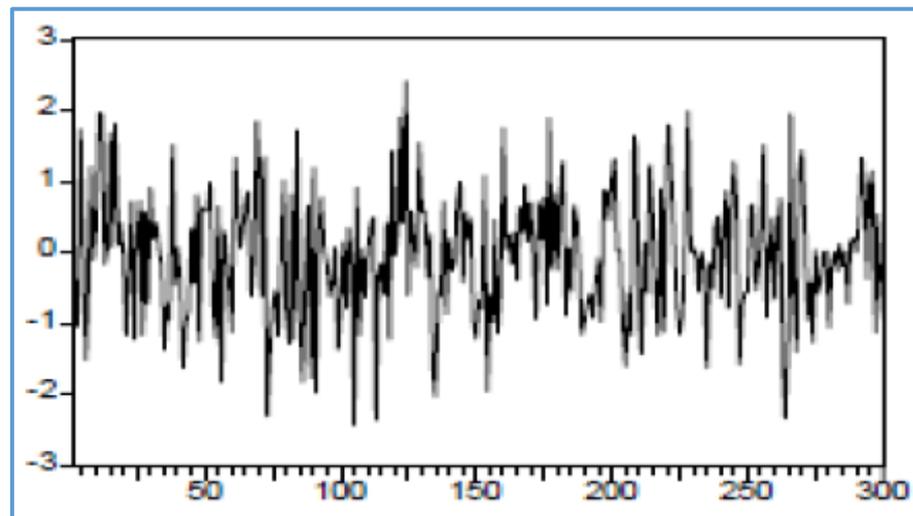
$$V(a_t) = 0 \quad \forall t$$

$$V(a_t) = E(a_t^2) = \sigma^2 \quad \forall t$$

$$\text{Cov}(a_t, a_t) = 0 \quad \forall t \neq S$$

Figura 1:
Proceso de un Ruido Blanco

Figura 3:
Proceso de un Ruido Blanco



2.2.7. Modelo univariante

Estos modelos univariantes son aquellos en los que sólo se observa una variable a la vez, y la serie temporal Y_t es un ejemplo de uno de estos modelos. Los modelos de este tipo se expresan en forma de polinomio. Entre los métodos unidimensionales, hay algunos métodos muy simples, tales como el modelo autorregresivo de primer orden, el modelo de tendencia lineal o exponencial, entre otros. (ChuquiHuayta, 2019, p. 18)

Modelos Univariantes No Integrado

Los Procesos Autorregresivos AR (p), las Medias Móviles MA (q) y Procesos Mixtos ARIMA (p, q), son considerados como modelos no integrados debido a que no intervienen al grado de diferenciación y la estacionalidad de la serie. (ChuquiHuayta, 2019, p. 18).

Modelo Univariante Integrado

Son aquellos modelos que se pueden obtener mediante suma o integración de un proceso estacionario; A estos modelos se les denomina también modelos no estacionarios homogéneos. (ChuquiHuayta, 2019, p. 19).

Series de Tiempo Estacionarias

Una serie de datos o valores que no muestra ninguna fluctuación sistemática en la media (la serie no indica una tendencia) ni variación en la varianza se conoce como serie estacionaria. En consecuencia, se considera que un proceso es estacionario si se produce en todos y cada uno de los puntos temporales de la serie.

En realidad, hay muchas series que no son estacionarias, pero la diferencia entre su primer y segundo orden sí lo es. El propósito de diferenciar la serie es



hacer que la serie sea diferencialmente estacionaria, pero recuerda que, si toman valores de la diferencia, también lo serán. estacionario, por lo que puede ocurrir una sobre diferenciación de la serie, ocasionando problemas de identificación del modelo que mejor representa el proceso seguido por el orden, y su varianza aumentará.

La serie temporal es de naturaleza estacional, ya que contiene fluctuaciones que se repiten periódicamente además de tendencias y ciclos a largo plazo. Por ejemplo, las observaciones mensuales pueden funcionar de manera similar a las observaciones del mismo mes; por ejemplo, las ventas de juguetes "en diciembre" también pueden tener patrones de comportamiento cíclicos que duran menos de un año, como "cada seis meses" a partir de junio. meses". Las observaciones de junio y diciembre tienen un comportamiento similar.

2.2.8. Componentes de una serie de tiempo

Los métodos de un estudio de valores de secuencias temporales exigen investigar para determinar los elementos que contribuyen a la variación de valor entre series individuales. Descomposición es el nombre que recibe el proceso de análisis, en el que cada componente se examina de forma aislada. A menudo es útil descomponer las series de tiempo en sus componentes principales. (Hanke, 2006).

Tendencia: cuando una serie tiene tendencia, no es raro que las observaciones sucesivas estén altamente correlacionadas, con coeficientes de correlación significativamente diferentes de cero y desapareciendo después de los primeros retrasos de tiempo.



Estacionalidad: Una serie se caracteriza por la estacionalidad, el efecto de un fenómeno que ocurre o se repite regularmente.

Aleatoriedad: si existe una correlación entre: y_t y y_{t-k} ; la serie es irregular o aleatoria; están cerca de 0 para cualquier desplazamiento de k . Los valores continuos de la serie temporal no están correlacionados.

Cíclico: Ocurre durante un largo período de tiempo, generalmente asociado con auges o recesiones económicas. Cuanto más largo es el ciclo, más difícil es identificarlo, principalmente porque no hay suficiente tiempo para recopilar información.

2.2.9. Modelos de series de tiempo

Operadores y Polinomios

Gutiérrez (2008). Los polinomios en capas son valiosos porque nos permiten prefigurar de manera exacta patrones muy valiosos (pero aparentemente complejos).

Operador de Retraso o BACKWARD B

Aplicable a z_t nos indica que se debe retrasar la variable un periodo:

Es decir:

$$B z_t = z_{t-1}$$

También:

$$B^2 z_t = B[Bz_t] = B[z_{t-1}] = z_{t-2}$$

Y en general:

$$B^k z_t = z_{t-k}$$

Operador diferencia ∇ aplicable a Z_t nos indica que debe obtener las diferencias entre Z_t y su valor rezagado:

$$\nabla Z_t = Z_t - Z_{t-1} = (1-B)Z_t$$

$$\nabla Z_t = \nabla(Z_t - Z_{t-1}) = (Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2})$$

Polinomios formados por observaciones presentes y pasadas ponderadas:

$$G(B) Z_t = Z_t - g_1 Z_{t-1} - g_2 Z_{t-2} - \dots - g_k Z_{t-k} = Z_t - \sum g_j Z_{t-j}$$

Polinomios de retraso racionales:

$$G(B) = A(B)/C(B)$$

$$A(B) = 1 - \sum a_j B^j ; C(B) = 1 - \sum c_j B^j$$

2.2.10. Modelo ARMA (p,q)

Un proceso aleatorio siguiendo una variable aleatoria Z_t ; su desviación del valor esperado μ viene dada por:

$$Z_t = Z_{t-\mu}$$

Expresamos el modelo de la siguiente manera: $\Phi(B)Z_t = \theta(B)a_t$

Donde $\Phi(B)$, $\theta(B)$; son operadores de retardo de orden p y q respectivamente, $\{a_t\}$, es una variante fortuita de un progreso de ruido blanco, media cero y varianza finita.

Otro método de redactar un progreso detrás de una voluble Z_t es:

$$(1 + \phi_1 B + \phi_1 B^2 + \dots + \phi_p B^p) Z_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) a_t \dots (i)$$

O también:

$$Z_t + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \dots \text{(ii)}$$

El modelo ARMA (p,q) es una extensión del modelo AR y del modelo MA que incluye las características de ambos modelos. Esta generalización se deriva del hecho de que las series temporales tienen atributos de los procesos AR y MA. Se ha visto que los procesos AR y MA presentan estas cualidades. La noción de robustez sugiere que, a la hora de desarrollar modelos, hay que utilizar el menor número de parámetros posible.

No se espera que todas las series temporales sean estacionarias, lo cual es una suposición al construir modelos ARMA; sin embargo, se sabe que, para alguna sucesión en estas condiciones, previamente, segunda o tercera diferencia a una sucesión es estacionalidad. considerado como un proceso inicial.

$\{Z_t\}$, Al sufrir de no estacionariedad debido a tendencias polinómicas no deterministas, denominada no estacionariedad homogénea, se pueden construir procesos estacionarios $\{w_t\}$, de tal modo:

$$w_t = \nabla^d Z_t \dots \text{(2)}$$

Para todos los t, se puede obtener un modelo ARMA para esta nueva serie:

$$\phi(B)w_t = \phi(B)a_t, \text{ equivalente a considerar el modelo ARIMA:}$$

$$\phi(B)\nabla^d Z_t = \phi(B)a_t$$

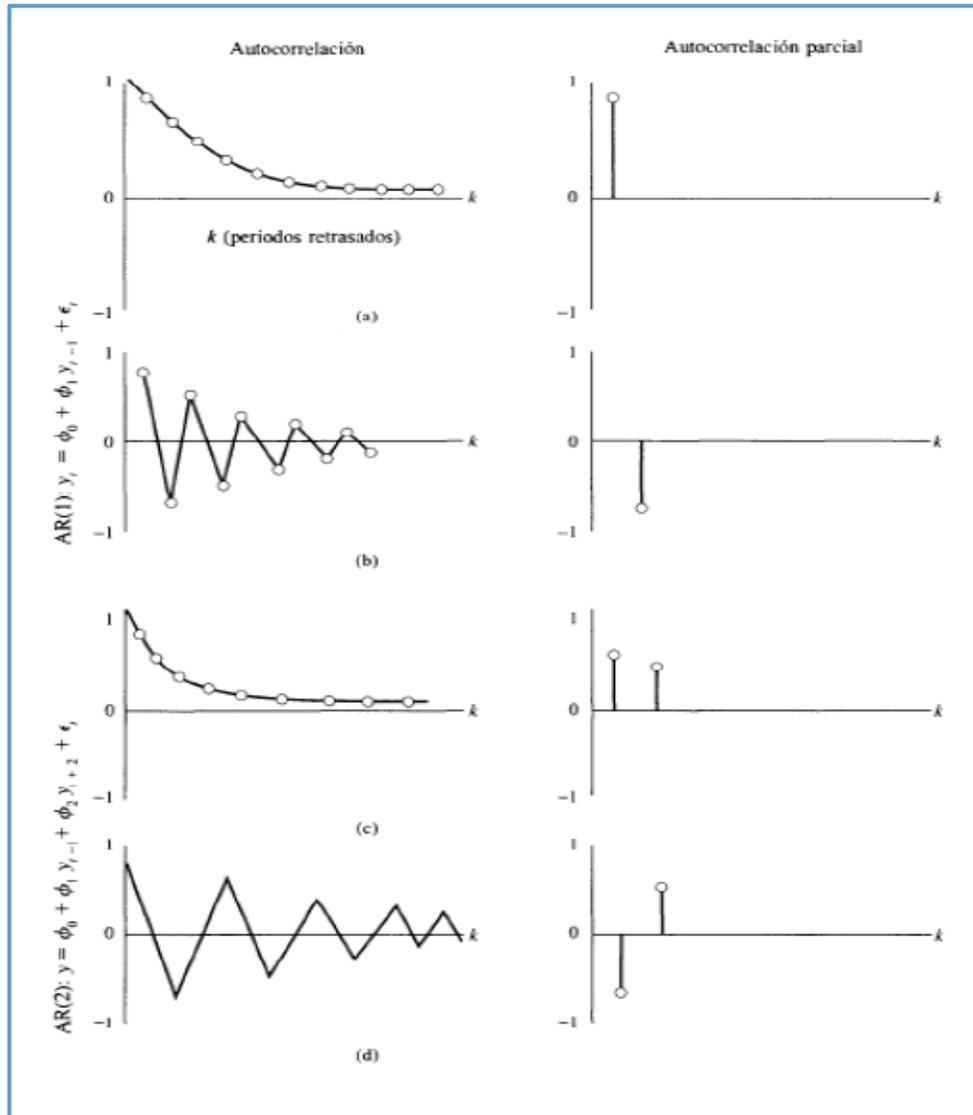
Modelo ARIMA: El término “integración” proviene de que Z_t , igual a un número interminable de datos contemporáneo y pretéritos de W_t , considerando la ecuación 2, para $d=1$; El valor de Z_t se puede lograr operando uno y otro, lados de la dicha ecuación por el operador ∇^{-1} , se obtiene:

$$S Z_t = \nabla^{-1} W_t = (1 - B)^{-1} W_t = W_t + W_{t-1} + W_{t-2} + W_{t-3} + \dots,$$

una suma de un número infinito de términos.

Figura 4:

Coefficientes de autocorrelación y autocorrelación parcial de los modelos ar(1) y ar(2)



$$AR(1) y_t = \phi_0 + \phi_1 y_{t-1} + E_t$$

$$AR(2) y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + E_t$$

Modelo Autorregresivo AR (1)

El caso más sencillo está representado por un modelo autorregresivo de primer orden en el que el valor del parámetro "C" se fija en 0. Un modelo autorregresivo de primer orden se define como;

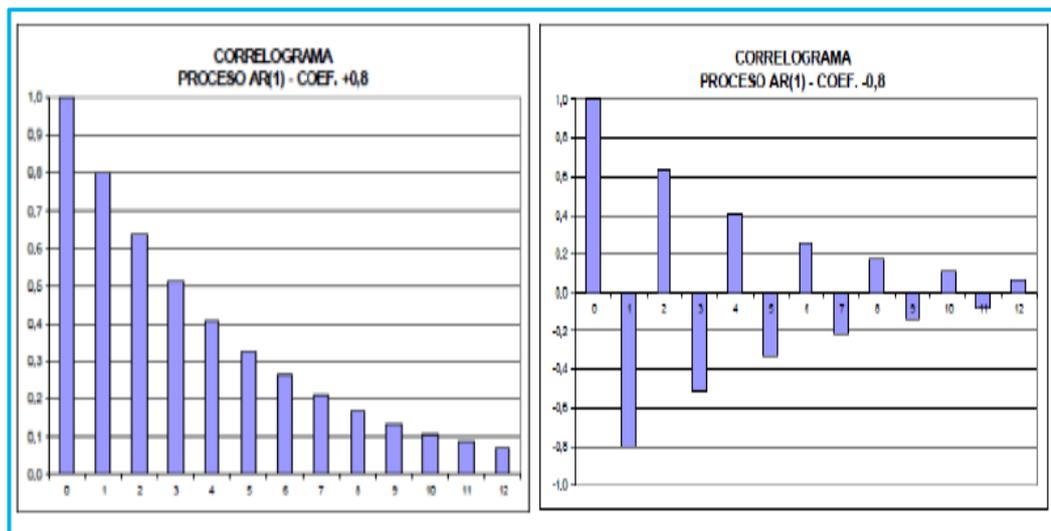
$$Y_t = \phi_1 Y_{t-1} + e_t \quad |\phi_1| < 1$$

Proceso AR (1)

Una norma AR (1) viene determinado por:

$$Y_t = \phi_1 Y_{t-1} + e_t$$

Figura 5:
Proceso AR (1) Correlograma



Proceso AR (2)

Un modelo AR (2) está definido por:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

Modelos de Medias Móviles

Los modelos de media móvil (MA) pronostican y_t basándose en una combinación lineal de errores anteriores de y_t , pero las formas progresivamente

(AR) se describen y_t como una facultad lineal de una serie de valores históricos reales de y_t . Los modelos autorregresivos (AR) son lo contrario de los modelos de media móvil (MA). Los modelos de medias móviles suelen indicarse con las iniciales MA, al igual que los modelos autorregresivos. Estos modelos se basan en términos independientes y errores múltiples convenientemente ponderados correspondientes a periodos pasados que explican el valor de una determinada variable en el periodo t . Los modelos de medias móviles son similares a los modelos autorregresivos en el sentido de que explican el valor de una determinada variable en el período t . Por lo tanto, el modelo con q términos de error se expresa como: $MA(q)$. (Chuquiwayta, 2019, p. 35).

Medias Móviles

El modelo de Box-Jenkins propone que las series de tiempo pueden explicarse mediante una combinación aleatoria de eventos que han sido renombrados en períodos anteriores.

Ningún fenómeno en la Tierra es un evento aleatorio. Por ejemplo, la venta de un producto se ve afectada por el lanzamiento de productos nuevos y diferentes; o el mercado de valores es bombardeado continuamente con nueva información aleatoria, cuanto más tiempo después del evento, menor impacto en la observación actual; como antes, su elección Coeficientes se estabiliza y no registra mayor poder explicativo más allá de este valor. Promedio móvil o proceso de orden q : $MA(q)$, se define de la siguiente manera:

$$y_t = \mu - a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

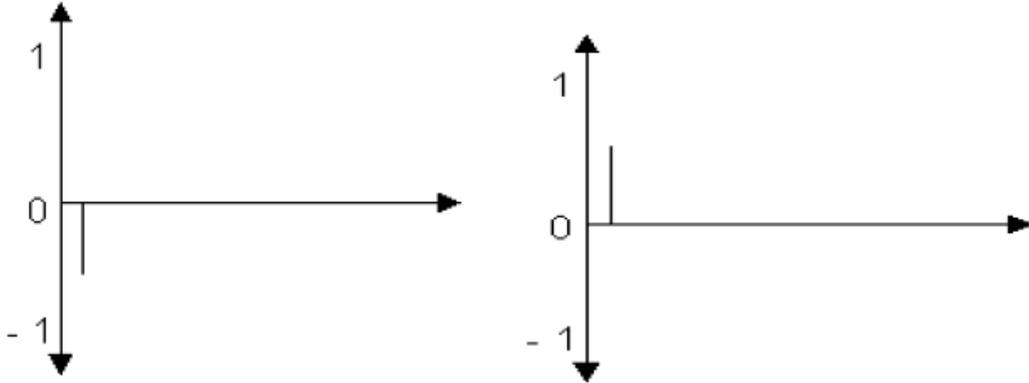
Donde a_t , es un ruido blanco con las propiedades ya definidas.

Proceso MA (1)

Una norma MA (1) viene determinado por:

$$y_t = \mu + e_t - \theta_1 e_{t-1}$$

Donde: e_t , un ruido blanco con las propiedades ya precisas.



Modelo MA (2)

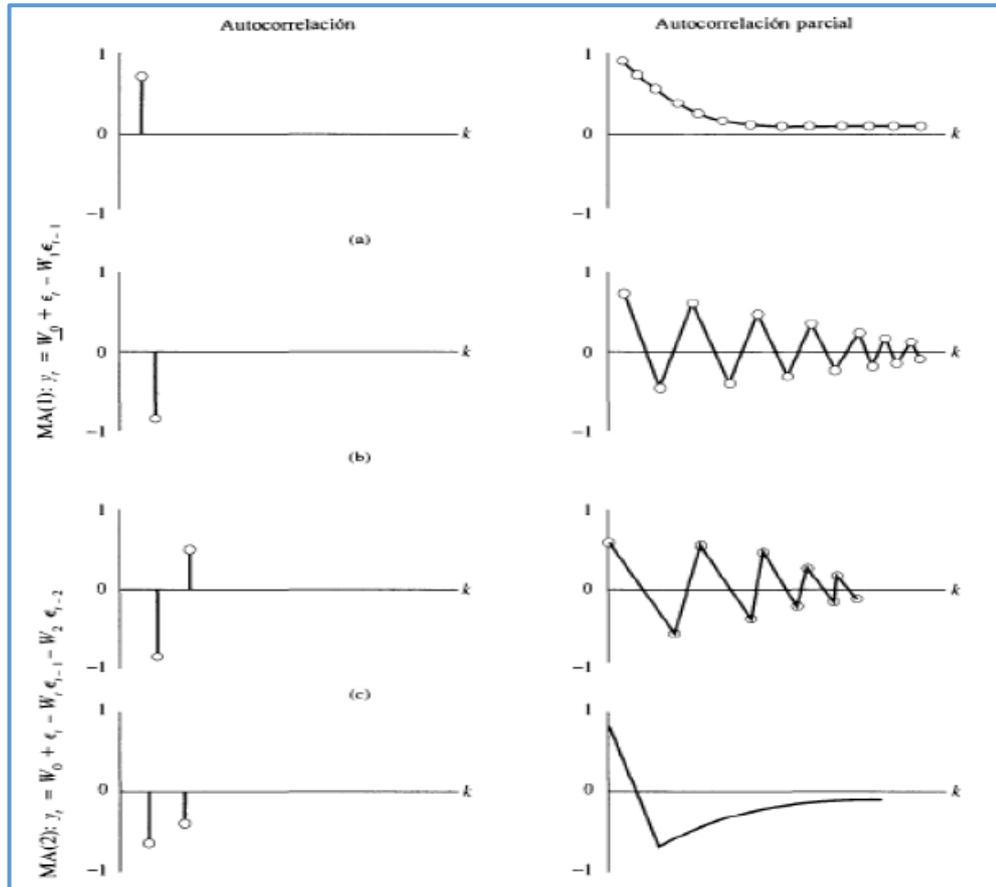
Una norma MA (2) suprimiendo una constante siendo así precisado por:

$$y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} = (1 - \theta_1 L - \theta_2 L^2) e_t$$

Donde: e_t , un ruido blanco con las propiedades ya precisas.

Figura 6:

Coefficientes de autocorrelación y autocorrelación parcial de los modelos MA(1) y MA (2).



$$MA(1): y_t = w_0 + \varepsilon_t - w_1\varepsilon_{t-1}$$

$$MA(2): y_t = w_0 + \varepsilon_t - w_1\varepsilon_{t-1} + w_2\varepsilon_{t-2}$$

Modelos mixtos autorregresivos de medias móviles (ARMA)

Un modelo ARMA se genera mediante la combinación de un modelo autorregresivo (AR) con un modelo de media móvil (MA). A continuación, se define el modelo ARMA (p, q):

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

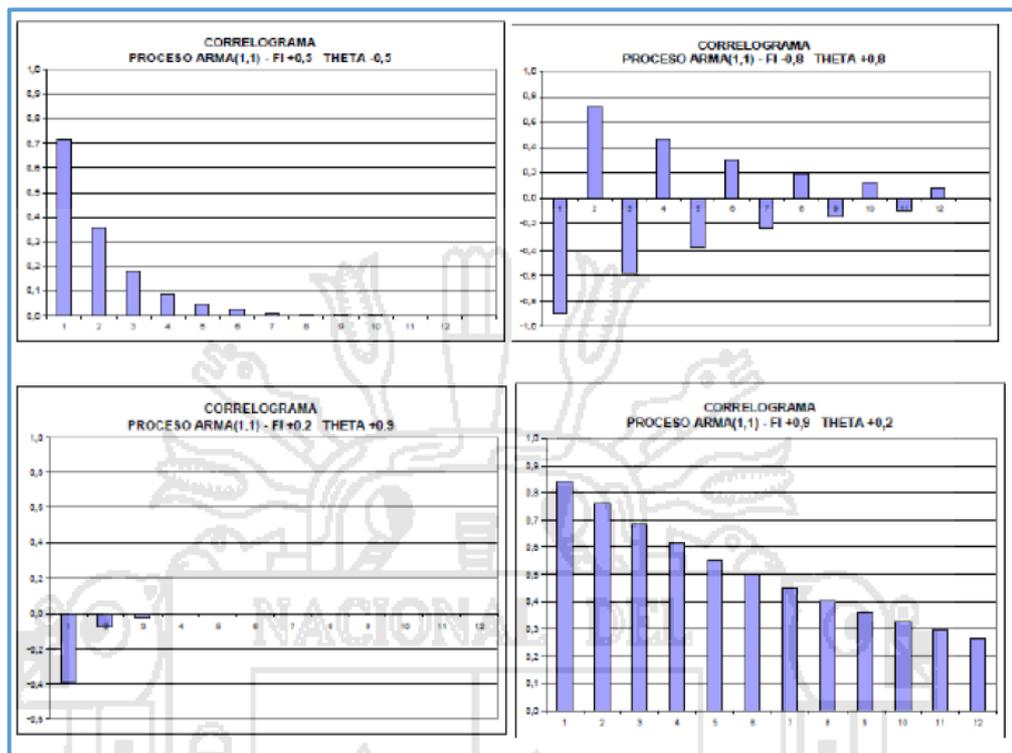
Modelos ARMA (1,1)

El análisis de Correlograma es especialmente importante para las series.

Para el proceso, el proceso ARMA (1,1), las constantes se omiten por simplicidad.

$$y_t = \varphi_1 y_{t-1} + a_t - \theta_1 a_{t-1}$$

Figura 7:
Correlograma Proceso ARMA (1,1)



Modelos lineales no estacionales

Modelo Arima (P,D,Q)

Dado que el proceso aleatorio que sigue a $[Z_{t-u}] = Z_t$ no es estacional en muchos casos, pero si es diferencial de primer orden, segundo orden, tercer orden... enésimo orden, ARMA generalización del modelo puede formularse para obtener un modelo ARIMA.

Queremos:

$$(B)[\nabla^d (Z_{t-u})] = (B) \nabla^d Z_t = \theta B a_t$$

El denominado modelo autorregresivo integrado y media móvil o modelo ARIMA (Autorregresive Integreted Moving Average). La norma ARIMA se detalla con mayor precisión como norma ARIMA (p,d,q), cual p es la cifra de retrasos ejecutados por el gestor de desplazamiento polinomial $\phi(B)$, procesa, d es el diferencial con respecto a Z_t , donde el gestor ∇^d , ejecución y q es la cifra de veces retrasadas por el gestor de desplazamiento polinomial $\theta(B)$.

La norma ARIMA (p,d,q) muestra que el patrón consiste en un polinomio progresivamente de regla p , orden d del diferencial de la variable estudiada Z_t , y un polinomio medio variable de orden q .

Los modelos ARIMA (p,d,q) con p : número específico de periodos de retraso, q : número de valores pasados y d : número de derivados; forman una clase especial de procesos no estacionarios. (Chuquiwayta, 2019, pág. 45).

Se define como:

$$W_t = \phi_0 - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} + \dots + \theta_2 Y_{t-p} - e_t - \phi_1 e_{t-1} - \phi_2 e_{t-2} - \dots \\ - \phi_2 e_{t-p}$$

Donde:

$$W_t = Y_t - Y_{t-1}, \text{ una diferencia.}$$

ϕ_i : número de retardos para la variable Y.

ϕ_1 : número de valores previos de residuo e_t .

e_t : error aleatorio ruido blanco.

Modelo Multiplicativo Estacional (ARIMA (P,D,Q) (p,d,q))

Combinación de efectos aleatorios estacionales y no estacionales influenciados por observaciones de ciertas características demográficas o series temporales.

BOX – JENKINS (1970): propuso una norma común de:

$$\phi(B^E) \nabla_E^D(z_{t-u}) = \theta(B^E) a_t$$

Donde estas volubles a_t , no se presumen ruido blanco, si no productores por una norma ARIMA (p,d,q), donde: $\phi(B)\nabla^d a_t = \theta(B)a_t$ con (a_t) , es un progreso de ruido blanco. Las dos por ultimar provisiones se alcanza el modelo multiplicativo estacional.

$$(B)\phi(B^E) \nabla_E^D(z_{t-u}) = \theta(B)\theta(B^E) a_t$$

La cual se demuestra por norma ARIMA (p,d,q).(P, D, Q)_E, de modo era de esperar, cuanto mayor sea la dificultad del modelo, más compleja será la estructura de autocorrelación correspondiente. Un modelo ARIMA que realiza una multiplicación estacional en una serie de observaciones mensuales nos permite tener en cuenta las posibles correlaciones entre años para las observaciones en el mismo mes. Es decir, capturando los efectos estacionales y de tendencia de los procesos "multiplicativos" o "auto-reforzadores" para lograr tales efectos. (Guerrero, 2003, p. 182).

2.2.11 Función de autocorrelación

En el campo del tratamiento de señales, una de las herramientas matemáticas que se utiliza a menudo se llama autocorrelación. Esta obligación de autocorrelación se determina así correlación cruzada de un signo consigo misma. Esta obligación de autorregresivo se usa para delimitar modelos reiterativos en las



señales, como la periodicidad de una señal oculta por el ruido, o para identificar una señal que no contiene un componente específico, sino que se produce en varias de sus frecuencias armónicas. Por ejemplo, la periodicidad de una señal oculta por ruido puede determinarse mediante la función de autocorrelación. En el campo de la estadística, la autocorrelación Y_t de una serie temporal discreta de un proceso se define como la correlación que existe entre ese proceso y una variación retardada de una misma serie temporal.

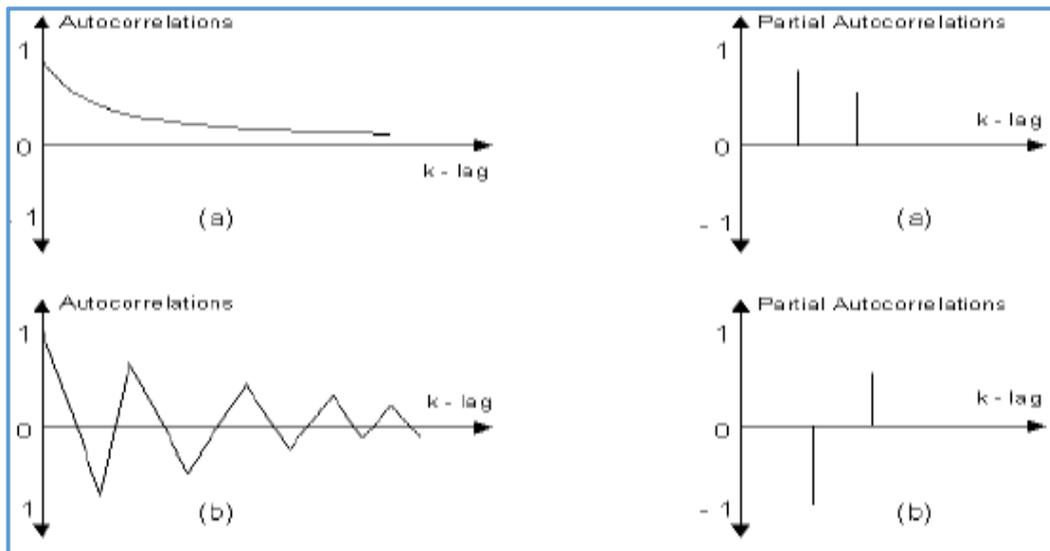
Si: Y_t denota el progreso estacionario de segundo orden cuyo dato inicial es μ , se denota como una función de autocorrelación.

$$R(K) = \frac{E[(Y_{j-\mu})(Y_{j-k} - \mu)]}{\sigma^2}$$

Donde:

E: es el valor esperado y k es el desplazamiento temporal considerando (generalmente desplazamiento). Esta función está en el rango $[-1,+1]$, donde +1 significa correlación perfecta (las señales se superponen perfectamente después de k lapsos de tiempo) y -1 significa anti-correlación impecable. Es un hábito frecuente en tantas doctrinas rechazar la normalización σ^2 y usar las conclusiones de autocorrelación y auto varianza indistintamente.

Figura 8:
Comportamiento de la función de Autocorrelación y Autocorrelación Parcial



2.2.12 Función de autocorrelación parcial (FACP)

Se utiliza para ayudar a determinar qué tan estrechamente se relaciona el valor real de una variable con los valores pasados dentro de ella. Los efectos de las demás variables siguen siendo los mismos. Función de autocorrelación parcial (FACP), del ejemplar P_{kk} , en el retroceso k es una correlación entre exploraciones (series de tiempo), que están apartadas k etapas de periodo. Manteniendo constantes las correlaciones en los rezagos intermedios es decir rezagos menores de k ; Entre otras palabras, la autocorrelación parcial es la correlación entre Y_t y Y_{t-k} , después de eliminar el efecto de las Y intermedias. (Chuquiwayta, 2019, p. 47).

2.2.13. Caminata al azar

El modelo de paseo aleatorio es simplemente un modelo AR (1) el proceso de caminata al azar se define como: (Gonzales, 2009).

$$y_t = y_{t-1} + a_t$$

En general: Dado la serie y_t , esto corresponde finalmente al logaritmo de los valores originales, que pueden ser representados por el proceso ARIMA (p,d,q) si su diferencia es del orden de "d". La letra I en ARIMA significa "Integración", lo contrario de diferenciación. Sí:

$$y_t = \nabla^d y_t, y, y_t; \text{ dado un proceso ARMA (p,q) estacionarios:}$$
$$(1 - \varphi_1 B^1 - \dots - \varphi_p B^p) Y_t = (1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q) a_t$$

y_t : Según el proceso ARIMA (p,d,q), la variable original y_t también se escribe como:

$$w_p(B)(1 - B)^d y_t = t(B) a_t$$

2.2.14. Transformación de box-cox

(**Box-Cox, 1964**) definió una transición instantánea se define de tal manera que varios períodos de tiempo más generales no participan simultáneamente en la transición:

$$y_t^\lambda = \begin{cases} (y_t^\lambda) / \lambda & \lambda \neq 0 \\ \lambda = 0 \end{cases}$$

En la transformación de Box-Cox, se debe definir el parámetro de transformación λ . Si el parámetro $\lambda=1$, la transformación de Box-Cox en realidad implica tomar el logaritmo; cuando el parámetro $\lambda=0$, se define como una ecuación cuadrática, una transformación logarítmica; la primera ecuación y, en el caso límite, el logaritmo de los números de nivel inicial también se cumplen.

2.2.15. Intervalos de confianza para las predicciones

La varianza de los errores en pronóstico se puede utilizar para realizar un intervalo de confianza para los pronósticos realizados mediante la siguiente expresión:

$$\Pr [y_{t+k} \pm \lambda_{\alpha} + \sigma_{et(k)}]$$

Si las innovaciones et , se asume una distribución normal, el parámetro λ_{α} , se recuperará de la tabla de dichas distribuciones para el nivel de significancia elegido α .

PREDICCIÓN DE UNA SERIE DE DIFERENCIA

Si un modelo ARIMA estima muchas diferencias, será preciso recuperar predicciones para un conjunto de diferencias; se pueden hacer diciendo que Y_t es la serie que nos interesa analizar.

$$E_t Y_{t+k} = E_t Y_{t-k} - E_t Y_{t+k+1}$$

Por lo que:

$$\begin{aligned} E_t Y_{t+k} &= E_t Y_{t-k} - E_t Y_{t+k+1} \\ &= Y_{t-k} + E_t Y_{t+k+1} + E_t Y_{t+k+2} + \dots + E_t Y_{t+k+1+et} \end{aligned}$$

ERROR DE PREDICCIÓN

La equivocación de una profetización de una diferencia entre la realización de una variable aleatoria y la profetización de ese dato; el error contenido en el pronóstico Y_{t+k} , depende del periodo en que dicha predicción se realiza.

Ezequiel (1985); El método de pronóstico de Box-Jenkins (1970) implica encontrar un modelo matemático que represente el comportamiento de una serie temporal de datos; como se mencionó anteriormente, el modelo utilizado en este trabajo es un modelo ARIMA unidimensional, donde el comportamiento de la serie de tiempo se basa en la secuencia misma. Las observaciones pasadas y los errores de pronóstico pasados, o la diferencia entre los valores reales pasados y

los valores actuales, se explican usando el modelo para hacer predicciones apropiadas. El modelo ARIMA tiene la siguiente estructura general:

$$\phi_p(B)(1-B)^d x_t = k + \theta_q(B)a_t$$

Donde:

x_t : se determinan la complementación en la época t de la serie objeto de estudio $\phi_p(B)$ y $\theta_q(B)a_t$, entre dos polinomios, de las órdenes de retardos.

$B(B_{\alpha_1} = X_{T-1})$, es la orden de las diferencias de primer orden que hay que tomar para realizar que una complementación sea estacionaria en media y a_t es una secuencia de ruido blanco con un componente estacional integrado en el modelo. La ventaja del modelo de predicción de Box-Jenkins es que, una vez que se gana experiencia con su enfoque, el mecanismo de búsqueda de patrones es más o menos rápido, gracias al uso de computadoras; además, una vez que se encuentra el modelo, se puede comparar inmediatamente con los datos reales para realizar predicciones y comparaciones. Predicciones de observaciones que pertenecen al pasado, por lo que es fácil ver gráficamente qué tan bueno o malo es el modelo elegido. Otra característica del modelo es que las mejoras a corto plazo se logran a largo plazo, lo que se debe principalmente a la construcción del modelo ARIMA. Todas estas conclusiones son generalizaciones, debido a que cada serie tiene sus propias características, para modelar series de tiempo con el método de Box-Jenkins, es necesario utilizar aplicaciones informáticas que faciliten la tarea, ya que, debido a la dificultad y gran cantidad de operaciones, es irrealizable sin la ayuda de una computadora. El método univariado de Box-Jenkins divide el proceso de modelado en cuatro fases.



A continuación, cada una de estas cuatro fases se explicará brevemente desde una perspectiva práctica sin entrar en demasiada explicación teórica más allá del alcance de este estudio.

IDENTIFICACION

Para identificar un patrón general, primero se deben analizar los cambios estacionales en la media y la varianza de la serie; de lo contrario, el diálogo es estático.

En la siguiente fase, si es necesario, determine el orden de diferenciación p y q de los polinomios autorregresivos $\theta_p(B)$, y de medias móviles $\theta_q(B)$, de la norma **ARIMA** la que más resaltan en el dibujo temporal de una serie y las funciones de autocorrelación (ACF), autocorrelación parcial (PACF).

Varios modelos alternativos pueden aparecer en esta etapa. En los modelos estacionales, es necesario probar si la serie es estacionaria en el componente estacional; en ese caso, tenga en cuenta las diferencias de pedidos estacionales.

VALIDACION

Una vez evaluado el modelo ARIMA identificado en el primer paso, se diagnostica su validez desde un punto de vista teórico. Hay muchas pruebas de diagnóstico y las más comúnmente aceptadas se han utilizado para refinar el estudio y se enumeran a continuación. Luego, primero se prueba la significación estadística de todos los parámetros calculados. A continuación, luego comprobamos si la serie temporal formada por los residuos del modelo, es decir, la diferencia entre el valor verdadero anterior de la serie y el valor predicho obtenido por el modelo, corresponde a la nula. analizar el ruido. Su ACF; es decir, ambos son válidos para la predicción, pero en este punto tiene más sentido



mantener solo uno de ellos, comparar los errores estándar (RSE) de todos los residuos y mantener el que posee el bajo valor.

PRONOSTICO

En este último periodo, se procesan previsiones con un modelo señalado al culminar de este periodo posterior. Por ello vuelve a ser primordial el uso del ordenador, señalando al programa el número de previsiones que se requieran para deducir y el pronóstico a partir del cual tiene que ser computadas.

2.6. DEFINICION DE TERMINOS BASICOS

SERIE

Es una sucesión de 'n' de exploraciones de valores dispuestas igualmente espaciadas en el tiempo ordenadas por función.

MODELO

Es un modelo de cualquier conjunto de supuestos bien definidos sobre las propiedades estadísticas de un proceso estocástico. (Mauricio, 2007, pág. 13).

MODELO BOX – JENKINS

El modelo es uno de los métodos de pronóstico basado en la estimación eficiente de parámetros en un proceso iterativo.

MODELO UNIVARIANTE DE BOX – JENKINS NO INTEGRADO

Estos progresos de Promedios Móviles (q), AR Autorregresivo (p) y Proceso Mixto ARMA (p,q), se consideran modelos no integrados porque no invierten la estacionalidad observada de la serie



MODELO UNIVARIANTE DE BOX – JENKINS INTEGRADO

Los progresos mixtos integrados norma ARIMA (p,d,q) Procesos Estacionales Mixtos Integrados ARIMA $(p,d,q)*(P,D,Q)$ Proceso una media móvil exponencial en el que interviene la estacionalidad de la serie ensayada.

RUIDO BLANCO

El proceso es completamente estocástico y las variables se distribuyen para tener media cero, varianza constante y sin autocorrelación entre las observaciones

ESTACIONARIEDAD

En la serie temporal, se dice que la serie es estacionaria si: $f(t) = f(t + k)$, es decir, el comportamiento de las variables en el tiempo es el mismo si la serie se desplaza,

ESTACIONALIDAD

Se define como la repetición regular de un determinado patrón de comportamiento.

Ejemplo: se puede repetir cada 3 meses, 6 meses, cada año, cada 4 años etc.

CORRELOGRAMA

Estos se trazan con valores separados de las funciones de autocorrelación total y parcial frente al retardo.

MODELO DE PRONOSTICO

Se define como una predicción, una declaración de lo que sucederá con los fenómenos físicos durante un período de tiempo.



CONSUMO DEL SECTOR DOMÉSTICO

La energía es la cantidad gastada en varios electrodomésticos utilizados en el hogar.

CONSUMO SECTOR COMERCIAL

El consumo en las empresas comerciales.

CONSUMO SECTOR INDUSTRIA

consumo empresarial.

CONSUMO GENERAL

Es la energía utilizada para iluminar calles, plazas, callejones y vías de uso público en las ciudades.

CENTRAL ELECTRICA

Unidades generadoras de electricidad incluyendo fuerza motriz y obras civiles necesarias.

SISTEMA DE ACCIONAMIENTO PRINCIPAL

Un sistema interconectado de generadores que permite el intercambio de energía eléctrica y la libre comercialización de la energía eléctrica.

USUARIO

Una persona que usa electricidad usando equipo con ciertas restricciones.

PROCESO

Un fenómeno que está en constante cambio a lo largo del tiempo.

ELECTRICIDAD



Este es el tipo de energía que se produce debido a la diferencia de potencial entre dos puntos. Esta situación permitirá la creación de una corriente entre dos puntos si se encuentran a través de un conductor eléctrico para lograr el trabajo anterior.

KILOWATT

Es una medida de la electricidad que se define como la cantidad de energía utilizada en un tiempo determinado. Esta unidad suele estar conectada a otras unidades más comunes, como la potencia de la célula de potencia, que es la cantidad de energía que pierde una célula y puede determinarse multiplicando su tensión por su corriente.

KILLOWATT - HORA

Es una unidad de trabajo o energía equivalente a producir un kilovatio de potencia en una hora, o 3,7 millones de julios, el símbolo es 'Kwh', que se usa para medir el gasto eléctrico.

MEGAVATIO - HORA

Un mega julio de electricidad equivale a un megavatio-hora de producción eléctrica. Se considera que un millón de vatios, a menudo denominados mil kilovatios, se suministran en una hora cuando se utiliza el prefijo métrico mega, que significa millón. La cantidad de energía que utilizan las grandes empresas urbanas o industriales se mide en megavatios hora (mwh). El mwh también se utiliza para determinar la capacidad de producción de las centrales eléctricas, mientras que la palabra megavatio-año, que se emplea para equiparar la energía anual suministrada por la central, también se utiliza en situaciones similares.



2.7. OPERACIONALIZACION DE VARIABLES

Tabla 1:

Operacionalización de variable.

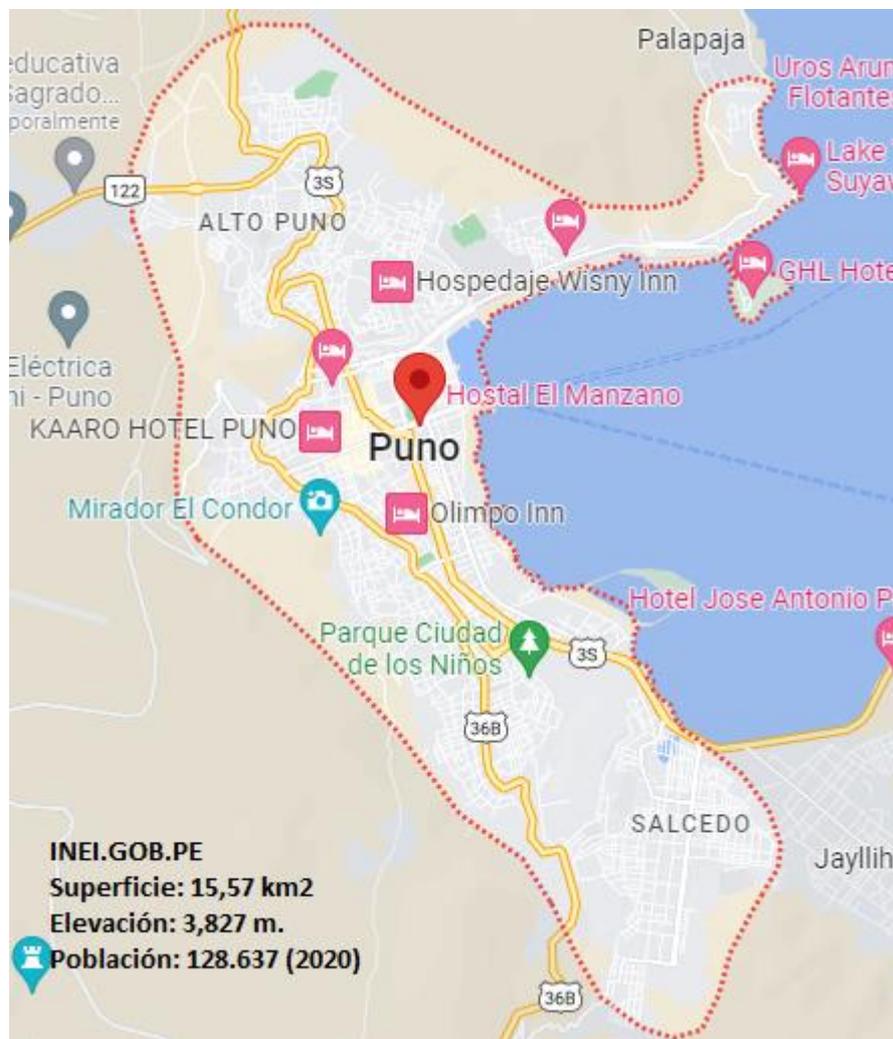
VARIABLES	INDICADOR	INDICE
Variable Dependiente: Consumo de energía doméstico (Mw-h/mes) megavatio por hora, en la zona residencial en la Ciudad de Puno, periodo 2011 – 2021.	Expresada en: Megavatio por hora.	Mw-h/h
Variable independiente: Consumo de energía doméstico (Mw-h/h) Megavatio por hora, desfasado en distintos periodos de tiempo en la zona residencial de la Ciudad de Puno.	Expresada en: Megavatio por hora	Mw-h/h

CAPITULO III

MATERIALES Y MÉTODOS

3.1. POBLACIÓN

Los registros actuales de uso de energía en las zonas residenciales de la ciudad de Puno que fueron recopilados por la Dirección de Planificación sirven como representación de la población considerada. (Electro Puno S.A.A.)



3.2. MUESTRA

La población se basa en el criterio del investigador y consiste en el consumo total de energía eléctrica local en la ciudad de Puno para el periodo del 2011 al 2021, se toma



en cuenta este criterio de selección por ser este periodo el más reciente y representativo elaborado por el planeamiento de oficina. Electro Puno S.A.A.

3.3. METODO DE RECOLECCIÓN DE DATOS

A efectos de esta investigación, los datos se han obtenido por métodos directos, utilizando la documentación facilitada por el servicio. de Electro Puno por la jefe de división de operaciones comerciales, Electro Puno S.A.A.

3.4. METODO DE TRATAMIENTOS DE DATOS

Este proyecto de indagación se empleará, teoría de WIENER KOLMOROV, también conocida como punto de vista de Box-Jenkins, para los periodos temporales.

A fin de usar los parámetros de predicción se fracciona los periodos temporales en los contiguos de comprobación y verificación. Se determinará como valores de preparación de enero 2011 hasta diciembre del 2021 siendo 132 datos y se deja desde diciembre del mismo año para en adelante la validación correspondiente de 12 datos.

Se realizará un análisis exploratorio de la serie y se aplica los modelos ARIMA, los pasos a seguir para obtener el modelo univariante mediante la técnica Box-Jenkins son los siguientes: Conceptualización de dibujo de la serie.

Operación de una función de autocorrelación (**FAA**) así como de la función de autocorrelación parcial (**FAAP**).

Técnica de identificación.

Estimación de los parámetros.

Proceso de verificación.

Técnica de elaboración de previsiones.



Luego de realizar las predicciones de los 12 valores futuros utilizando las dos técnicas de previsión se opera las medidas de error, el coeficiente y se procesa la demostración luego conocer la técnica que presenta los pronósticos más precisos. Por terminar se realiza el identificado que técnica presenta previsiones más precisas que obtuvieron el error porcentual absoluto medio de las 12 pronósticos de dicha técnica, para validar el límite de periodo en donde los pronósticos son excelentes.

La investigación es:

. Según el método de investigación es cuantitativa debido a que la variable a analizar es el índice de (MW-h/mes Megavatio por hora) y los valores representa el consumo diario energía doméstico del índice.

Por lo tanto, el objetivo es aplicada con la finalidad del estudio, modelar y pronosticar usando métodos de predicción.

Por lo tanto, el nivel de ahondamiento en el objeto de investigación es predictiva ya que se requiere lograr un pronóstico al futuro empleando métodos estadísticos, es este caso se realiza las normas ARIMA.

Por lo tanto, el tratamiento de variables es no experimental para este trabajo de investigación los datos se obtuvieron de las oficinas de Electro Puno S.A.C. esta plataforma tiene un registro diario del consumo de energía eléctrica doméstico de la región Puno.

Por lo tanto, la inferencia es inducir por que se pretende realizar pronosticaciones de acuerdo a los métodos que proporcionan menor error en el ajuste del modelo.

Por lo tanto, la etapa temporal es longitudinal debido a que los valores estén tomados en el periodo.



METODO DE BOX – JENKINS – TEORIA DE WIENER KOLMOROV.

(Ezequiel, 1985). El enfoque utiliza un proceso que se divide en cuatro etapas, cada una de las cuales consta de las etapas siguientes:

Uno de los modelos predictivos que utiliza métodos iterativos para llegar a estimaciones precisas de los parámetros del modelo se conoce como enfoque Box-Jenkins.

- Responsabilidades que recaen sobre los hombros del analista.
- Acciones que lleva a cabo un ordenador.

En pocas palabras, para que la técnica Box-Jenkins pueda predecir con exactitud el desarrollo futuro de una acción de una acción, es importante validar que este modelo de comportamiento no se ha alterado a lo largo del tiempo.

PRONOSTICO

En primer lugar, ARIMA crea la serie temporal de interés, a continuación, estima los parámetros del modelo ARIMA adecuado y, tras una fase de validación, utiliza este modelo para realizar predicciones con el menor número posible de errores de previsión. Una vez concluida la etapa de identificación, ARIMA genera las series temporales de interés.

FUNCION DE AUTOCORRELACIÓN

Función que se produce por las correlaciones internas entre los componentes de un periodo observado (gasto eléctrico total de los hogares) en la ciudad de Puno en el periodo 2011 al 2021.

Está definida por:

$$r(k) \frac{\text{cov}(y_t, Y_{t-k})}{r(0)} = \frac{E(y_{t-u})(Y_{t-k} - u)}{r(0)}$$

Donde:

$r(0)$: Es la auto covariancia cuando no existe desplazamiento alguno, es la varianza del proceso a la que se ajusta al consumo de energía eléctrica doméstico.

u : Es la media del proceso a la que se ajustan la serie de consumo de energía eléctrica doméstico.

$\text{cov}(y_t, Y_{t-k})$: Es la covarianza de la serie original y la serie desplazada en k periodos.

FUNCION DE AUTOCORRELACION PARCIAL

La matriz de autocorrelación para la serie estacionaria de longitud N , está dado por:

$$P_N = \begin{bmatrix} 1 & r_1 & r_2 & \dots & r_{N-1} \\ r_1 & 1 & r_1 & \dots & r_{N-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{N-1} & r_{N-2} & r_{N-3} & \dots & 1 \end{bmatrix}$$

En conjunto de autocorrelaciones parciales en varios desplazamientos, están definidos por:

$$\phi_{kk} = \frac{|Q_k|}{|P_k|}$$

Donde:

$|P_k|$: Es la determinante de la matriz de autocorrelaciones de orden $k.k$.



$|Q_k|$: Es la determinante de la matriz de autocorrelaciones. Con la última columna reemplaza a las funciones de autocorrelación generada por la serie de consumo de energía eléctrica doméstico.

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix}$$

r_k : La k-ésima función de autocorrelación del proceso a la que se ajusta la serie de consumo de energía eléctrica.

N = tamaño de la serie conformado 3650 días equivalentes a 10 años (2011 – 2021) de la serie original.

CONSTRUCCION DE MODELOS ESTOCASTICOS

Se pueden presentar dos casos durante la preparación o construcción del modelo, que se generan a partir del proceso de identificación de series temporales o generación de series. La identificación del modelo se realiza de forma iterativa, a través de filas que conducen a la validación. La norma ARIMA (*construcción p,d,q*) consta de las siguientes periodos:

- Identificación.
- Estimación
- Inspección o diagnóstico.

En este método, la fase de identificación de la construcción del modelo es la más importante, y la construcción del modelo en sí es una cuestión de inferencia estadística. Esto significa que, dada una colección de observaciones de una serie temporal, tiene que ser capaz de generar un modelo que revele el comportamiento a priori. El comportamiento



en sí tiene que confirmarse a priori, y el cumplimiento de las condiciones es una etapa más del proceso. El "análisis exploratorio de datos"

FASE DE IDENTIFICACIÓN DE MODELOS ESTOCÁSTICOS

Es de una disposición de estacionariedad de una serie (d y λ) que a continuación el número de parámetros autorregresivos (p) con media móvil (q), Viene hacer, si el modelo promediado en el tiempo es una serie no estacionaria, entonces se aplican las transformaciones adecuadas para convertirlos en estacionarios e invertibles, indicando el nivel de diferenciación y el algoritmo de Box-Jenkins, que utiliza más. Una representación gráfica de la serie muestra la fluctuación relativa a la media para corroborar la estacionariedad de los periodos.

La consideración de una función de autocorrelación, así como la función de autocorrelación parcial, se corroborará la significancia de los, r_k y ϕ_{kk} ; para la aceptación que ninguno de los parámetros estimados sea superior a 1 ni menor que -1 .

Calcula la raíz de la característica ecuación y comprueba la estacionariedad de la serie en el proceso de identificación. Esta prueba se puede realizar convenientemente solo cuando la raíz está en el círculo unitario.

FASE DE VERIFICACION DEL MODELO

- Para especificar un modelo ARIMA, se buscará el modelo que mejor represente el comportamiento de la secuencia y cumpla con los siguientes requisitos:

- Los residuos del modelo estimado corresponden aproximadamente a un comportamiento de "ruido blanco".

- El modelo calculado es estacionario y viceversa.



- Los coeficientes son estadísticamente significativos y están poco correlacionados entre sí.

- Los coeficientes del modelo son suficientes para representar el rango entre ellos.

- Alto grado de ajuste frente a otros modelos alternativos.

FASE DE PREDICCIÓN O PRONÓSTICOS

- Una vez que se encuentra un modelo adecuado, se puede hacer el pronóstico eligiendo un período de origen diferente.

- A medida que haya más datos disponibles, se puede construir el mismo modelo para el pronóstico eligiendo un período de inicio diferente.

- Si la serie parece cambiar con el tiempo, es necesario calcular los parámetros o incluso desarrollar un nuevo modelo.

Para predecir los diferentes patrones, tenemos:

$$y_t = B + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p}$$

$B, \phi_1, \phi_2, \dots, \phi_p$: Estimación de los parámetros que intervienen en la previsión.

P : Representa el número de periodos venideros para los que k es menor que cero.

y_{p+k} : Una media móvil, un estado mixto o un estado estacionario son los tres resultados posibles que puede predecir el pronóstico de un proceso.

MODELOS MIXTOS INTEGRADOS ARIMA (p,d,q)

PROCESOS ARIMA – NO ESTACIONARIOS

Los requisitos de estacionariedad y/o invertibilidad se impusieron a los procesos o modelos que se determinaron manejar en la fase anterior a ésta; los resultados de esta

determinación se denominan generadores de procesos no estacionarios. Según Box y Jenkins, así se define un modelo ARIMA:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t^\lambda = (1 - \theta_1 B - \dots - \theta_q B^q) a_t$$

Donde:

d : Es el número de diferencias necesarias para alcanzar la estacionariedad.

$|\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q|$: Se trata, por orden, de los coeficientes del componente autorregresivo, de la media móvil y de la propia media móvil.

B : Es el operador de retardos.

λ : Es el parámetro de la transformación Box – Cox.

$\phi_p(L)$: Es el operador polinomial del proceso de media móvil invertible, es decir las raíces de $\theta_q(B) = 0$ se caen fuera del círculo unitario.

a_t : Se afirma que las desviaciones tienen media cero y varianza constante a lo largo del tiempo cuando hay ruido blanco. El ruido blanco es la serie de desviaciones idénticamente distribuidas y no correlacionadas. Sin embargo, es interesante explorar sólo algunos tipos de no estacionariedad que son adecuados para describir el comportamiento de las series de consumo y que pueden convertirse en procesos estacionarios. Esto se debe a que sólo estas pocas formas de no estacionariedad son útiles para explicar el comportamiento de las series de consumo.

El proceso integrado x_t se desarrolló un proceso estacionario que tiene en cuenta el orden (q), y el proceso autorregresivo integrado de media móvil que se empleó se denominó proceso autorregresivo integrado de media móvil ARIMA (p,d,q):

$$\mathbf{AR}(p) = \mathbf{ARIMA}(p,0,0) \cong \mathbf{ARIMA}(1,0,0)$$

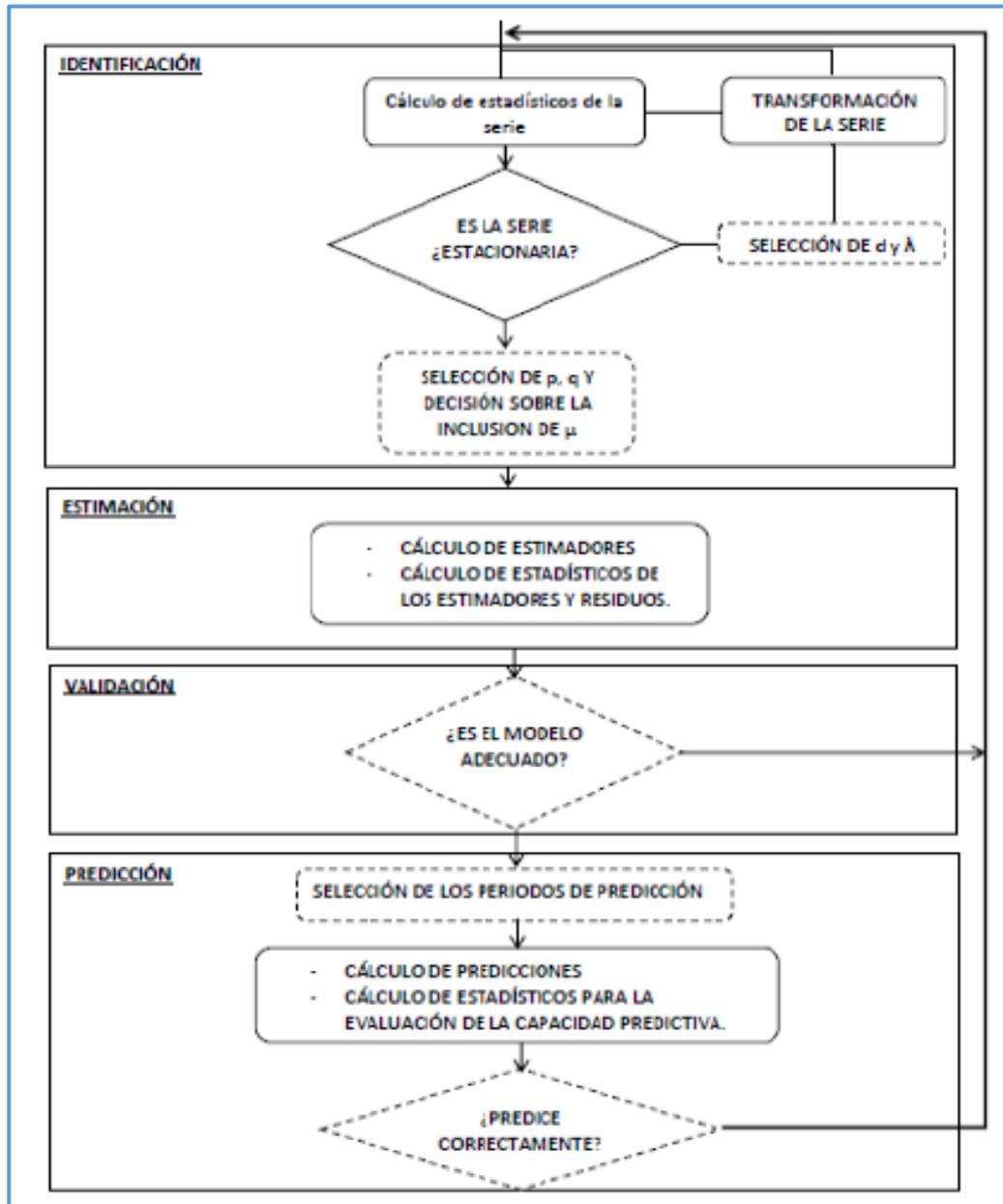
$$\mathbf{MA}(q) = \mathbf{ARIMA}(0,0,q) \cong \mathbf{ARIMA}(0,0,1)$$



$$\mathbf{ARIMA(p,q) = ARIMA(p,0,q) \cong ARIMA(1,0,1)}$$

Esto demuestra que los modelos ARIMA pertenecen a la categoría de procesos no estacionarios, y es fácil eliminar cualquier sesgo desconocido en los valores calculando las diferencias de primer orden entre ellos.

Figura 9:
Metodología del enfoque BOX-JENKINS





CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. APLICACIONES DE LA METODOLOGÍA BOX – JENKINS

En este estudio hemos utilizado como punto de partida este enfoque, que incluye cuatro pasos básicos: identificación, estimación, validación y predicción, con el objetivo de identificar un modelo univariado que mejor se ajuste a la serie histórica. consumo eléctrico doméstico.

Por lo tanto, las tablas y los gráficos se utilizan para analizar, discutir e interpretar datos.

Los datos originales correspondientes a la serie de electricidad domiciliaria son los siguientes. (Mw-h/mes Megavatio por hora) para el servicio zona residencial de PUNO – Electro Puno, Periodo 2011 – 2021.



Tabla 2:

Series históricos del consumo de energía doméstico eléctrico

(Mw-h/mes Megavatio por hora) en la ciudad de Puno, periodo 2011-2021.

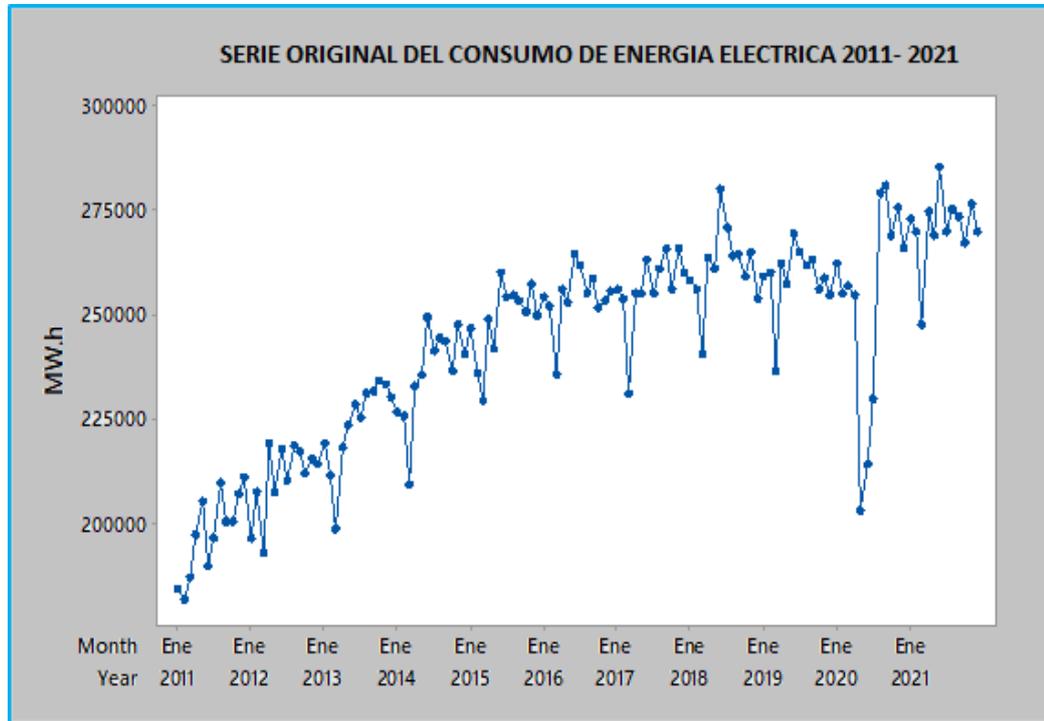
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
ENERO	184718	196729	219142	226770	247049	254528	256191	258321	259343	262289	273145
FEBRERO	181842	207522	211875	225744	236261	252217	253765	256225	260141	255341	269844
MARZO	187394	193149	198698	209598	229535	235841	231274	240747	236435	257272	247726
ABRIL	197488	219285	218353	232964	249170	256126	255255	263965	262371	254952	275003
MAYO	205400	207725	223630	235562	242050	253245	255143	260885	257344	203420	269108
JUNIO	190050	217971	228678	249353	260381	264635	263390	280176	269454	214384	285501
JULIO	196726	210427	225559	241380	254428	261909	255241	270759	264898	230113	269972
AGOSTO	209726	218625	231395	244463	254866	255500	260935	264125	261786	279244	275426
SETIEMBRE	200542	217612	231900	243555	253487	258678	266185	264578	263414	281047	273364
OCTUBRE	200542	212320	234584	236475	250767	251708	256163	259512	256069	269218	267183
NOVIEMBRE	207163	215742	233653	247553	257562	253332	266030	264905	259063	275954	276849
DICIEMBRE	211094	214302	230530	240498	249919	255723	260292	254021	254808	265948	269955

FUENTE: ELECTRO PUNO S.A.A.

4.1.1. Identificación del modelo

Figura 10:

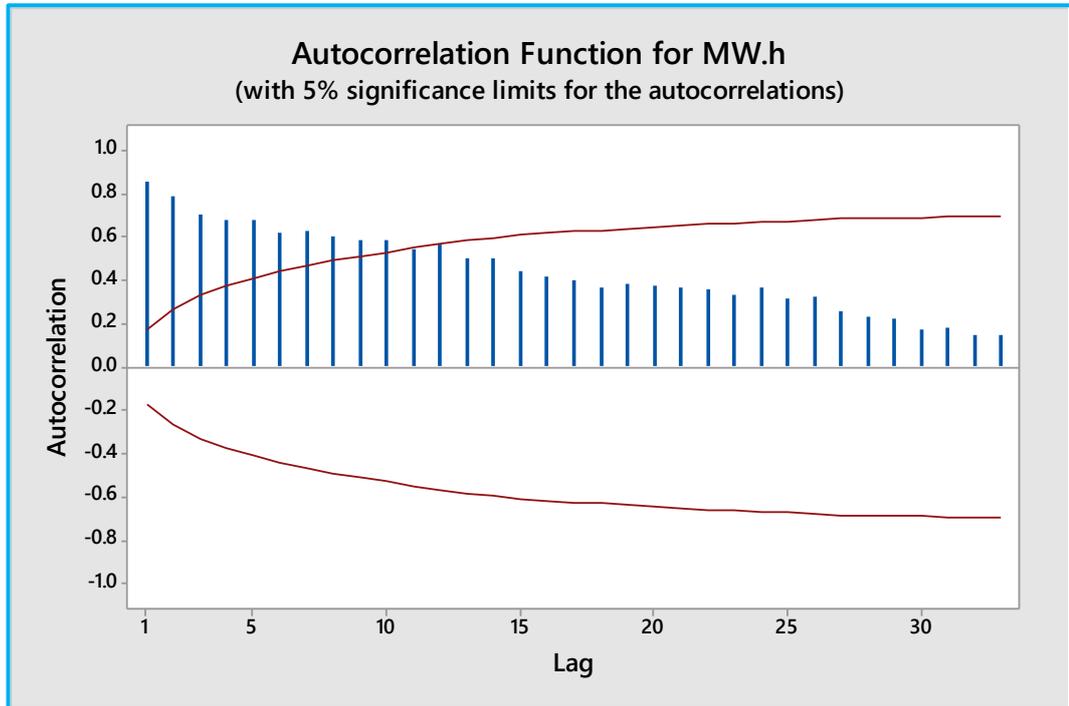
Consumo de energía eléctrica de la ciudad de Puno, Periodo 2011 -2021



En la figura N°10, Nos muestra los valores mensuales de una serie de consumo energético domiciliario en la ciudad de Puno, donde se aprecia un comportamiento muy variable, además no se mueven alrededor de la media ni de la varianza del proceso, lo que demuestra que, según los datos brutos, las series están al mismo nivel que la encuesta. No es estable. El gráfico muestra que esta serie tiene una tendencia creciente, por lo que ciertamente es claro que el consumo de electricidad de los hogares en esta serie no es estacionario.

Figura 11:

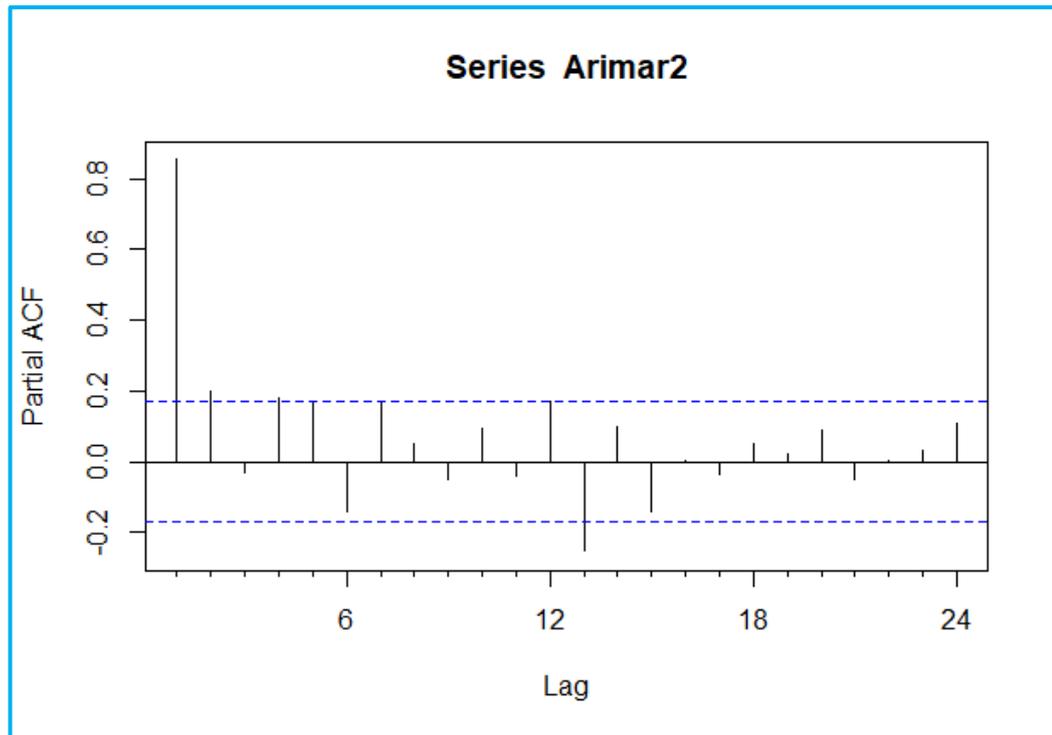
Función de autocorrelación estimada de la serie consumo de energía eléctrica doméstico de la Ciudad de Puno



En la figura N°11, Se muestra que el comportamiento de la función de autocorrelación disminuye, por lo que la autocorrelación calculada tiene un decaimiento aritmético lento, y se puede observar que el primer tiempo de retardo es significativamente diferente de cero, y luego el valor disminuye gradualmente dentro del intervalo de confianza. Entonces se dice que los valores están en tendencia y debemos tratar de diferenciarlos para formar una serie estacionaria.

Figura 12:

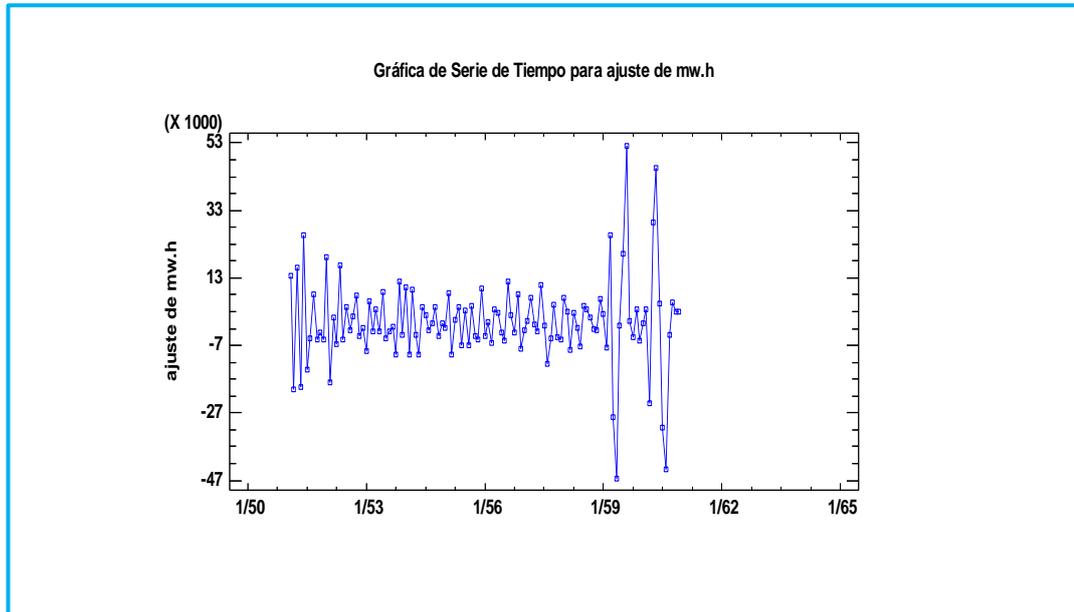
Función de autocorrelaciones parciales estimada de la serie consumo de energía eléctrica doméstico de la Ciudad de Puno.



En la figura N°12, Se puede observar que tiene un coeficiente de autocorrelación significativo y es el primer coeficiente del que podemos estar seguros que es diferente de cero, esta secuencia tiende a cero a partir del segundo, también tiene inversión de signo, lo que demuestra que la serie no es estacionaria y se acerca a un ruido blanco.

Figura 13:

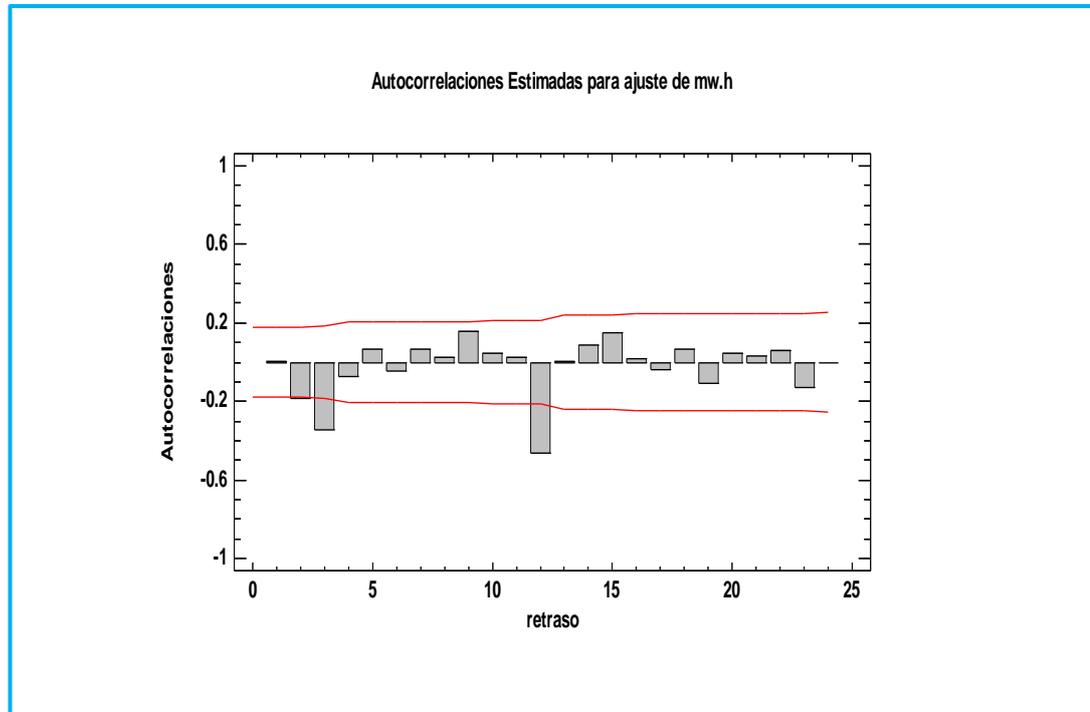
Función de dos diferencias no estacional por una diferencia estacional para la serie consumo de energía eléctrica doméstico de la Ciudad de Puno.



En la figura N°13, Como podemos ver, el gráfico muestra una trayectoria en el tiempo y se puede ver que no hay señales de una tendencia en la serie, y también se puede observar que algunos valores muestran picos que difieren entre meses y años. En tal caso procedemos a expresar que estamos frente a una serie de valores casi estacionarios lo cual fue comprobado en las autocorrelaciones.

Figura 14:

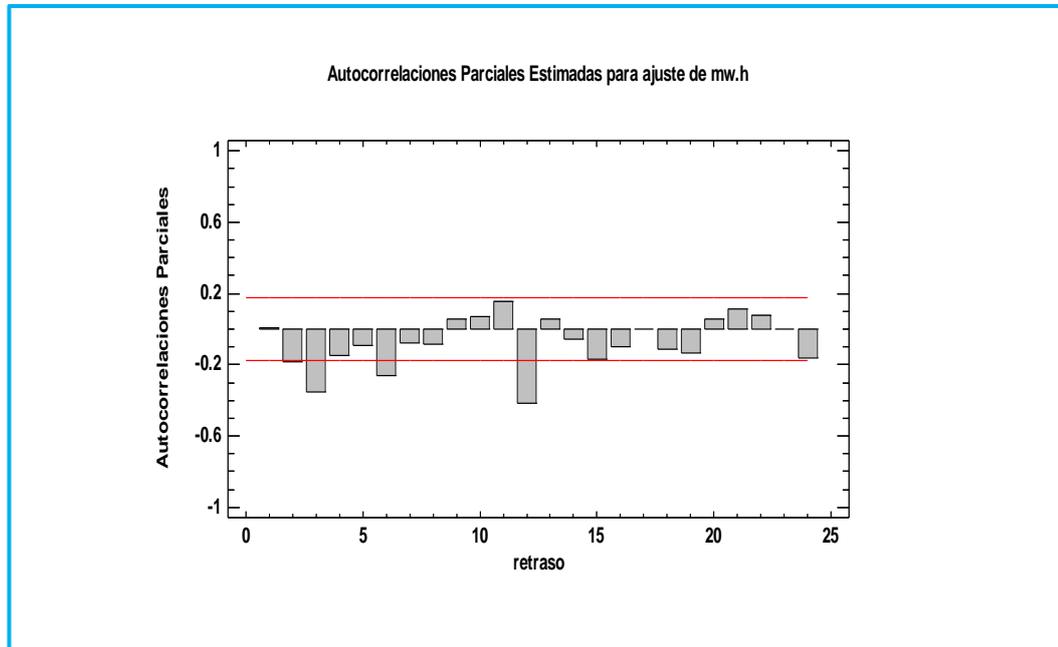
Función de autocorrelaciones estimadas para dos diferencias no estacional por una estacional para la serie consumo de energía eléctrica doméstico de la Ciudad de Puno



En la figura N°14, Se puede observar que el valor de la primera diferencia no estacional calculó la función de autocorrelación en la serie histórica de consumo eléctrico doméstico en la ciudad de Puno. El patrón de tendencia ha desaparecido, después del primer cambio significativo, los cambios 4 y 12 fuera del intervalo de confianza son significativos, el resto está dentro del intervalo permitido y diligentemente dentro del intermedio de confianza, procedemos a expresar que los valores ya no son significativos.

Figura 15:

Función de autocorrelaciones parciales estimadas para dos diferencias no estacional y una estacional para la serie de consumo de energía eléctrica doméstico de la Ciudad de Puno.



En la figura N°15, Los valores presentan un declive exponencial resaltándonos la idea de que es un modelo de media móvil y SMA móvil, y la función de autocorrelación parcial estimada nos sugiere un proceso integrado ARIMA (2,1,1)x(0,1,1) multiplicativo que procederá a detallar la conducta de la serie de consumo energía eléctrica y aun no es estacionario.

MODELO TENTATIVO IDENTIFICADO

Todos los análisis basados en gráficos de autocorrelación y funciones de autocorrelación parcial y diferencias seriales identificaron modelos alternativos, los probaron mediante el método Box-Jenkins y seleccionaron buenos modelos para la predicción.

Concluimos que el modelo determinado para la serie histórica de consumo eléctrico domiciliario en la ciudad de Puno es el modelo multiplicativo ARIMA (2, 1, 1) (0, 1, 1)₁₂ con ecuaciones de la siguiente forma:

$$\hat{Y}_t = 2Y_{t-1} - Y_{t-2} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_{12}\varepsilon_{t-12} + \theta_{13}\varepsilon_{t-13}$$

4.1.2. Estimación del modelo identificado

Después del proceso de identificación del modelo, estimaremos el modelo ARIMA (2,1,1)(0,1,1)₁₂ sobre los valores de la serie de consumo de energía de los hogares.

Resumen de Pronósticos

Diferenciación no estacional de orden: 1

Diferenciación estacional de orden: 1

Modelo de pronóstico seleccionado: ARIMA(2,1,1)x(0,1,1)₁₂

Número de pronósticos generados: 12

Número de periodos retenidos para validación: 0

<i>Estadístico</i>	<i>Periodo de Estimación</i>	<i>Periodo de Validación</i>
RMSE	7751.97	
MAE	4690.32	
MAPE	1.9278	
ME	-163.937	
MPE	-0.118357	

Seguidamente, presentamos los cuadros de análisis de varianza para los modelos estimados.

Resumen de Modelo ARIMA

Tabla 3:

Resumen del modelo arima

<i>Parámetro</i>	<i>Estimado</i>	<i>Error Estd.</i>	<i>t</i>	<i>Valor-P</i>
AR (1)	0.637072	0.107513	5.92556	0.000000
AR (2)	-0.221441	0.0980305	-2.2589	0.025775
MA (1)	0.851407	0.0661172	12.8772	0.000000
SMA (1)	0.904686	0.0263639	34.3153	0.000000

FUENTE: ELECTRO PUNO S.A.A.

ELABORADO: EJECUTOR DE LA INVESTIGACION

La salida resume la significancia estadística de los términos en el modelo de pronósticos. Términos con valores-P menores que 0.05 son estadísticamente diferentes de 0 con un nivel de confianza del 95.0%. El valor-P para el término AR (2) es menor que 0.05, de modo que es estadísticamente diferente al valor 0. El valor-P para el término MA (1) es menor que 0.05, de modo que es estadísticamente diferente a 0. El valor-P para el término SMA(1) es menor que 0.05, de modo que es estadísticamente diferente a 0. El ruido blanco de entrada tiene una desviación típica estimada de 87205.

Cada estadística se basa en los errores de previsión de una sola entrada, que son las disparidades entre los datos observados en el momento t y el valor previsto en el momento $t-1$. Los tres estadísticos iniciales cuantifican el alcance de los defectos. Un modelo superior arrojaría un valor menor. El sesgo se mide

con los dos últimos estadísticos. Un modelo superior produciría valores más próximos a cero.

ESTIMACION DE LOS PARAMETROS DEL MEJOR MODELO DE PRONOSTICO SELECCIONADO: ARIMA (2, 1, 1) (0, 1, 1)¹²

Estimación de los parámetros del modelo identificado, ARIMA (2, 1, 1) (0, 1, 1)¹² para el consumo de energía eléctrica doméstico de la Ciudad de Puno, periodo 2011 – 2021

Ecuación del pronóstico Modelo multiplicativo ARIMA (2, 1, 1) (0, 1, 1)¹²

$$\hat{Y}_t = 2Y_{t-1} - Y_{t-2} + Y_{t-12} - Y_{t-13} - (0.851407)\varepsilon_{t-1} - (0.904686)\varepsilon_{t-12} + 0.770255\varepsilon_{t-13}$$

Como muestra la estimación de los parámetros para el modelo que corresponde y que es propuesto para la serie histórica de consumo de energía eléctrica doméstico de la Ciudad de Puno, periodo 2011 – 2021.

4.1.3. Validación o adecuación del modelo

Una vez seleccionado el modelo ARIMA multiplicativo y estimados sus parámetros, es necesario determinar si el modelo seleccionado se ajusta suficientemente a la serie histórica de consumo eléctrico de los hogares. Los detalles son para ver si los residuos del modelo estimado cumplen con la propiedad de validación, es decir corresponden aproximadamente al comportamiento del ruido blanco.

$$|\theta_1| < 1 \quad |\theta_{12}| < 1$$

$$\text{MA}(1) \quad \theta_1 = 0.851407 < 1$$



$$\text{SMA}(12) \theta_{12} = 0.904686 < 1$$

El p estadístico < 0.05

$$P = 0.00000 < 0.05 \text{ MA}(1)$$

$$P = 0.00000 < 0.05 \text{ SMA}(12)$$

CONTRASTE GLOBAL DE BOX Y PIERCE

Prueba de chi cuadrado ARIMA(2,1,1)x(0,1,1)₁₂

Ho: Los residuos sigue un proceso ruido blanco: $p_k > 0$

H1: Los residuos no siguen un proceso ruido blanco: $p_k > 0$

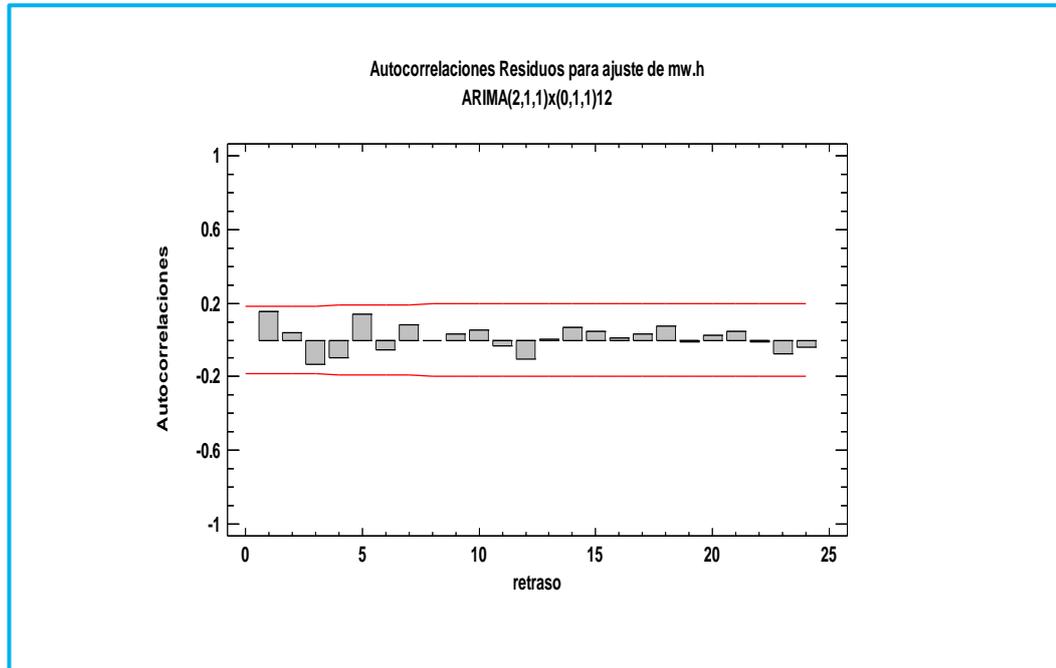
Niveles de significancia: $\alpha = 0.05 = 5\%$

Prueba estadística: $Q_{cal} = 0.803966$

Por consiguiente, como $Q_{cal} = 0.803966 < X^2_{(21,0.05)} = 14.5056$ se procede a aceptar la Ho y rechazar la H1, por consiguiente, se deduce que los residuos siguen un procedimiento de ruido blanco. Después de expresarse que la serie histórica de consumo mensual de energía eléctrica doméstica de la ciudad de Puno es estacionaria.

Figura 16:

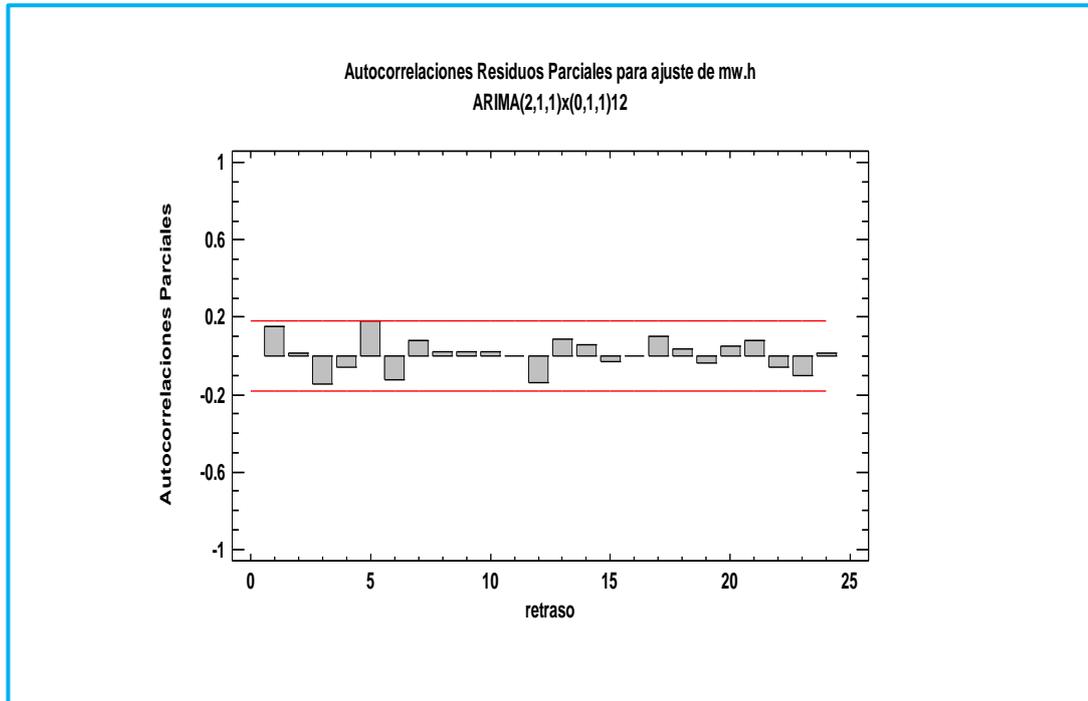
Función de autocorrelación de residuales estimadas para la serie histórica del consumo de energía eléctrica doméstico de la Ciudad de Puno.



En la figura N°16, Se ha demostrado que la serie se encuentra en algún punto cercano al intervalo de confianza. Como resultado, se puede concluir que las cifras de la serie sobre el uso de energía en los hogares son completamente arbitrarias.

Figura 17:

Función de autocorrelación parcial de residuales estimados para la serie histórica de consumo de energía eléctrica doméstico de la Ciudad de Puno.



En la figura N°17, Dado que la distribución de la función de autocorrelación parcial mostró que ningún coeficiente es estadísticamente significativo, de este resultado puede extraerse la conclusión de que la serie es aleatoria.

4.1.4. Fase de predicción

PRONOSTICOS ESTIMADOS

Tabla 4:

Datos de predicción estimadas (Mw-h/mes)

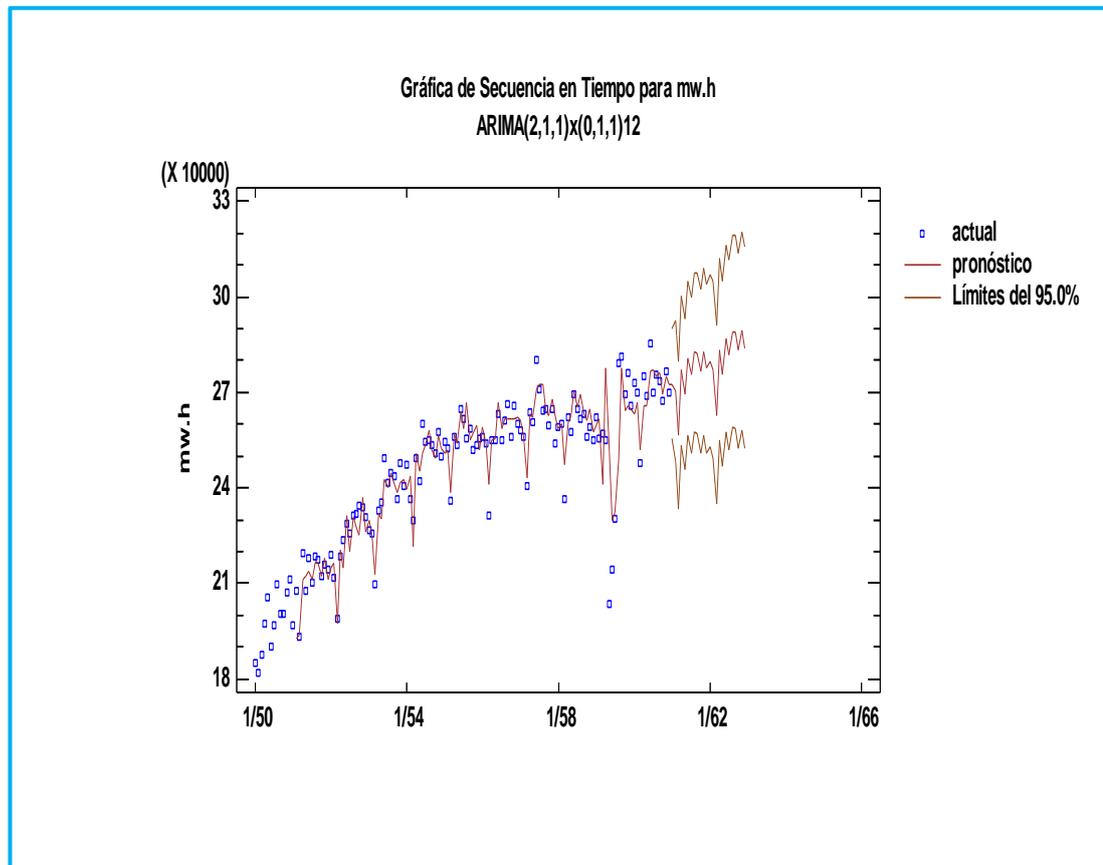
		Límite en 95.0%	Límite en 95.0%
<i>Periodo / mes</i>	<i>Pronóstico</i>	<i>Inferior</i>	<i>Superior</i>
1	272656.	255382.	289930.
2	270491.	248524.	292458.
3	256412.	233235.	279588.
4	276950.	253384.	300516.
5	269348.	245501.	293195.
6	280630.	256459.	304801.
7	275420.	250874.	299966.
8	282663.	257723.	307603.
9	282464.	257133.	307795.
10	276622.	250910.	302335.
11	282886.	256800.	308972.
12	277447.	250995.	303900.
13	279870.	252737.	307002.
14	277205.	249483.	304927.
15	262869.	234682.	291055.
16	283354.	254762.	311946.
17	275776.	246793.	304758.
18	287084.	257712.	316456.
19	281886.	252124.	311647.
20	289130.	258982.	319278.
21	288929.	258399.	319460.
22	283086.	252179.	313994.
23	289350.	258070.	320629.
24	283911.	252264.	315558.

En la tabla N°4, En la tabla muestra los valores pronosticados para MW.h. Los datos y residuos previstos del modelo ajustado (datos-previsión) se presentan para todo el tiempo del que se dispone de valores. Esto incluye tanto el principio como el final del periodo. Las limitaciones de las predicciones con un nivel de

confianza del 95,0% se muestran para los periodos de tiempo que se extienden más allá de la serie temporal. En caso de que el modelo ajustado a los datos sea adecuado para los valores, estos límites sugieren, con una probabilidad del 95,0%, dónde puede estar el valor genuino de los datos en el momento futuro designado.

Figura 18:

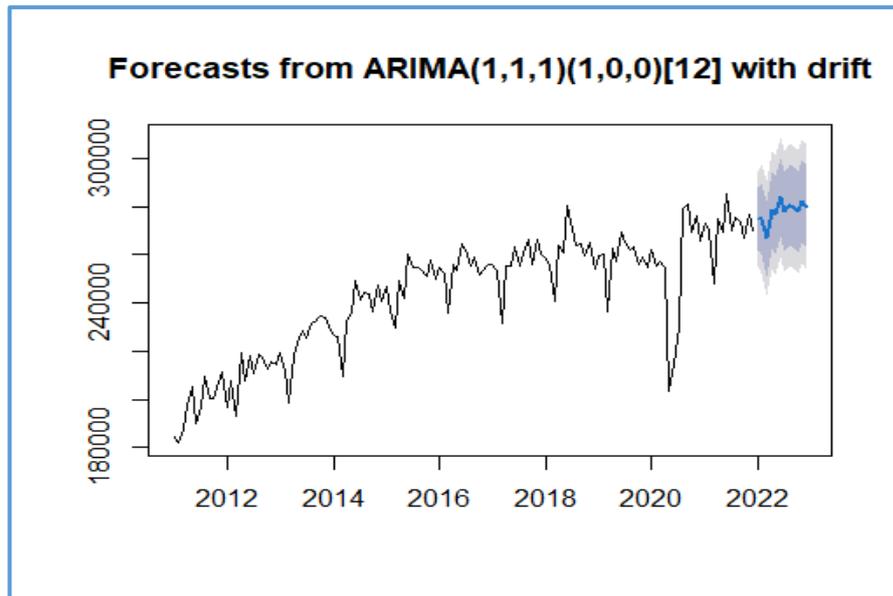
Función de pronóstico con un 95% de confianza para la serie de consumo de energía eléctrica doméstico de la ciudad de Puno, periodo 2011 – 2021.



En la figura N°18, El pronóstico nos muestra una tendencia creciente, un comportamiento que explica el aumento paulatino de la población, y el estándar del paquete estadístico nos lo muestra, y en un período observado de 12 (1 año), dicho pronóstico sigue evolucionando al mismo tiempo. dirección tierra adentro. Serie histórica de consumo de electricidad. En este sentido, podemos decir que el modelo resultante es bueno para el propósito que proponemos.

Figura 19:

Función de pronóstico con límite de confianza 95% de consumo de energía eléctrica doméstico periodo 2011 – 2021.



En la figura N°19, se observa los pronósticos e intervalo de confianza que nos demuestra por defecto y son para 12 periodos de tiempo donde resaltan, que dichos pronósticos siguen en el mismo sentido de la serie histórica de consumo de energía doméstico, en tal sentido se puede expresar que el modelo conseguido es bueno para nuestras intenciones proyectadas.

4.3. DISCUSIÓN

Serie de tiempo es un método de regresión que se ha venido siendo utilizando para analizar el comportamiento el pronóstico, debido a sus capacidades de adaptabilidad, generalización, aprendizaje y relaciones no lineales.

Estos modelos ARIMA como técnicas paramétricas y el modelo se entrena a través de regresiones lineales.

Estos modelos ARIMA se exponen patrones de forma conducta como tendencia y ciclos o estacionalidades y aleatoriedad.



Este modelo ARIMA es sensible a valores atípicos convergiendo al modelo de ruido blanco. Entonces el modelo ARIMA con los valores arreglaran el error generado por los valores atípicos convergiendo al modelo exacto.

En los modelos ARIMA se denominan parámetros a los valores a estimar.

Comparación de modelos

Variables de datos: mw.h

Número de observaciones = 132

Índice Inicial = 1/50

Intervalo de Muestra = 1.0 mes(es)

Longitud de la estacionalidad = 12

Modelos

ARIMA(2,1,1)x(0,1,1)₁₂

FORECASTS ARIMA(1,1,1)x(1,0,0)₁₂

(C) Media constante = 243095 Ajuste estacional: Multiplicativo.

(D) Promedio móvil simple de 3 términos Ajuste estacional Multiplicativo.

(E) Suavización exponencial simple con alfa = 0.0982



Ajuste matemático:

<i>Modelo</i>	<i>RMSE</i>	<i>MAE</i>	<i>MAPE</i>	<i>ME</i>	<i>MPE</i>
(A)	7751.97	4690.32	1.9278	-163.937	-0.118357
(B)	24863.7	20149.6	8.69605	14.1254	-1.04277
(C)	10562.5	5667.28	2.34909	1190.63	0.404159
(D)	8862.72	5112.51	2.17383	671.465	0.23548
(E)	11948.5	7902.14	3.31319	383.171	0.0225165

<i>Modelo</i>	<i>RMSE</i>	<i>RUNS</i>	<i>RUNM</i>	<i>AUTO</i>	<i>MEDIA</i>	<i>VAR</i>
(A)	7751.97	OK	OK	OK	*	***
(B)	24863.7	OK	***	***	***	***
(C)	10562.5	***	*	***	OK	***
(D)	8862.72	OK	**	**	OK	***
(E)	11948.5	OK	***	***	OK	***

ARIMA (1,1,1)(1,0,0)[12] with drift

Coefficientes :					
	ar1	ma1	sar1	drift	
	0.5335	-9.9341	0.4097	620.5467	
s.e.	0.0959	0.0409	0.0856	214.4873	
sigma^2	97972906	Log likelihood	-1390.47		
AIC	2790.95	AICc	2791.43	BICC	2805.32

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
156.3246	9708.852	6683.341	-0.01889128	2.809368	0.5848817	0.03485457

$$|\theta_1| < 1 \quad |\theta_{12}| < 1$$

$$MA(1) \quad \theta_1 = -9.9341 < 1$$

$$SMA(12) \quad \theta_{12} = 0.4097 < 1$$

RMSE = Root Mean Squared Error (Raíz del Cuadrado Medio del Error)



RUNS = Prueba corridas excesivas arriba y abajo.

RUNM = Prueba corridas excesivas arriba y debajo de la mediana.

AUTO = Prueba de Box-Pierce para autocorrelación excesiva.

MEDIA = Prueba para diferencia en medias entre la 1 mitad y la 2 mitad.

VAR = Prueba para diferencia en varianza entre la 1 mitad y la 2 mitad.

OK = no significativo ($p > 0.05$)

* = marginalmente significativo ($0.01 < p \leq 0.05$)

** = significativo ($0.001 < p \leq 0.01$)

*** = altamente significativo ($p \leq 0.001$)

Dado los valores, se puede observar los resultados de cinco diferentes modelos de pronósticos, viendo las estadísticas del error, el modelo con la menor raíz cuadrada del error cuadrado medio (RMSE) durante el periodo de estimación del modelo es el modelo (A), el modelo con el menor error medio absoluto (MAE) es el modelo (A). El modelo con el menor porcentaje del error medio absoluto (MAPE) es el modelo (A). También resume los resultados de cinco pruebas para determinar si cada modelo es adecuado para los datos. Un “OK” significa que el modelo pasa la prueba, un “*” significa que no pasa la prueba al nivel de confianza del 95%, dos “**” significa que no pasa la prueba al nivel de confianza del 99%, tres “***” significa que no pasa la prueba al nivel de confianza del 99.9%. Se puede notar que el modelo actual seleccionado, el modelo (A), pasa las tres pruebas.

Con los resultados obtenidos se concluye que las predicciones realizadas con el modelo multiplicativo $ARIMA(2,1,1) \times (0,1,2)_{12}$ son más precisos.



V. CONCLUSIONES

- 1) Los modelos univariantes integrados nos suministraron un buen ajuste para la serie consumo de energía eléctrica para el servicio eléctrico de la ciudad de Puno.
- 2) La serie consumo de energía eléctrica en el servicio eléctrico de la ciudad de Puno, muestran una inclinación progresiva y nos demuestra signos de variaciones cíclicas estacionales.
- 3) El óptimo modelo univariado para detallar y pronosticar los patrones de consumo de electricidad de los hogares en Puno es el siguiente modelo multiplicativo:

ARIMA (2, 1, 1) (0, 1, 1)₁₂ modelo multiplicativo

$$\hat{Y}_t = 2Y_{t-1} - Y_{t-2} + Y_{t-12} - Y_{t-13} - (0.851407)\varepsilon_{t-1} - (0.904686)\varepsilon_{t-12} \\ + 0.770255\varepsilon_{t-13}$$



VI. RECOMENDACIONES

- 1) Electro Puno S.A.A. La empresa - Puno, no facilita la adquisición de información sin ningún tipo de restricciones previo documento y coordinación para realizar estudios periódicos, incluyendo otras variables como: inversiones, personal, etc. Esto permite una mejor comprensión del presupuesto de la empresa, infraestructura, equipamiento técnico y otras variables.
- 2) Se sugiere realizar estos tipos de trabajo de investigación cada lapso de tiempo con el fin de conseguir frutos actualizados y que nos admita tener una superior visión de los residentes.
- 3) Prevenir la sobre-parametrización y la sobre-diferenciación, por ello concluyen a la producción de modelos erróneos.
- 4) Se recomienda el proceso de estimación usar las herramientas necesarias para comprobar la estacionariedad e invertibilidad del proceso, como son los de Dickey-Fuller, Box-Pierce, Ljung-Box, entre otros.



VII. REFERENCIAS BIBLIOGRÁFICAS

- COAQUIRA, R. J. (2018.).** Modelo Univariante para pronosticar la cantidad de ventas mensuales de bolsas de cemento rumi producidos en la planta cemento sur del distrito de Caracoto. Caracoto.
- CORDOVA, Z. (2006).** Estadística Inferencial. . Lima Perú: Talleres gráficos.
- GUERRERO, C. (2009).** Análisis estadístico de series de tiempo econométrico. Mexico: THONSON.
- GUTIERREZ. (2008).** Series de Tiempo. Editorial Mc GranH.
- MAMANI, P. C. (2017).** Modelo univariante para el consumo mensual de energía doméstica en el Distrito de Putina – Electro Puno. Putina.
- MESA, M. M. (1946).** Tópicos de Estadística Descriptiva y Probabilidad . Thales S.R. Ltda. .
- OLAYA, P. C. (2003).** Estudio del pronóstico de la demanda de energía eléctrica, utilizando modelo de Series de Tiempo Scientia et Technica Año IX, No 23. Pereira: Universidad Tecnológica .
- Osinergmin. (2018.).** ENCUESTA RESIDENCIAL DE CONSUMO Y USOS DE ENERGÍA. Lima: GPAE OSINERGMIN.
- Soria, G. (2010).** CONSUMO DE ENERGIA ELECTRICA PARA USO DOMESTICO EN SAN JUAN DEL RIO QUERETARO. Distrito Federal, Mexico: Universidad Autónoma Metropolitana Unidad Azcapotzaico.
- URIEL. (1985).** Análisis de series de tiempo ARIMA . Editorial Paraninfo.



ARANFO, F.O. et al. “Pronóstico de los índices accionarios DAX y S&P 500 con redes neuronales diferenciales”. Contaduría y Administración [en línea]. 2013, (México) 58(3), pp. 203-225, [Consulta: 20 diciembre 2019]. ISSN. 0186-1042. Disponible en: [http://dx.doi.org/10.1016/S0186-1042\(13\)71227-0](http://dx.doi.org/10.1016/S0186-1042(13)71227-0)

BOEHMKE, B. & GREENWELL, B. Hands-On Machine Learning with R [en línea] [Consulta: 13 diciembre 2019]. Disponible en: <https://bradleyboehmke.github.io/HOML/index.html#conversions-used-in-this-book>

CABEZON MANCHADO, M. Predicción demanda eléctrica española. Implementación de redes neuronales en Python [en línea] (Tesis) (Maestría) Universidad Copulense de Madrid, España. 2018. [Consulta: 10 diciembre 2019]. Disponible en: <https://eprints.ucm.es/49444/1/2018-MIGUEL%20CABEZON%20Memoria.pdf>

CALVO, D. Red Neuronal Recurrente – RNN. Diego Calvo [blog]. [Consulta: 12 diciembre 2019]. Disponible en: <http://www.diegocalvo.es/red-neural-recurrente/>

FLOREZ, R. & FERNANDEZ, J. Las Redes Neuronales Artificiales Fundamentos teóricos y aplicaciones prácticas [en línea]. S.l.: s.n. ISBN 9788497452465. Disponible en: <https://books.google.com.ec/books?id=X0uLwi1Ap4QC&pg=PA56&lpg=PA56&dq=red+maquina+de+cauchy&source=bl&ots=gLMzelpr3j&sig=7bOm9fYQj6gwP6zZs&hl=es#v=onepage&q=redmaquinadecauchy&f=true>

GRAPH. Machine Learning: Qué es, tipos, ejemplos y cómo implementarlo. [blog]. [Consulta: 19 de enero 2020]. Disponible en:



<https://www.grapheverywhere.com/machine-learning-que-es-tipos-ejemplos-y-como-implementarlo/>

KREPRI, WILLIAM. Redes Neuronales ¿qué son? [blog]. Disponible en:

<https://medium.com/@williamkhepri/redes-neuronales-que-son-a64d022298e0>

REYES LORA, F. Predicción de la demandad eléctrica: comparativa arima-redes neuronales mediante software SPSS. [en línea] (Tesis) (Maestría), pp. 123.

Escuela Superior de Ingenieros. Disponible en:

<https://idus.us.es/xmlui/bitstream/handle/11441/48443/Fco.ManuelReyesLora-ProyectoFindeMáster.pdf?sequence=1>

VILLAREAL M.M. Estudios de pronósticos en series de tiempo: Evaluación de métodos estadísticos y selección del mejor métodos con redes neuronales, tesis (licenciatura), México, Autónoma de Nuevo León, Nuevo león, 2006



ANEXOS



Tabla 5:

Valores de predicción estimadas de consumo de energía eléctrica domestico

(Mwh/mes)

		<i>Límite en 95.0%</i>	<i>Límite en 95.0%</i>
<i>Periodo / mes</i>	<i>Pronóstico</i>	<i>Inferior</i>	<i>Superior</i>
1	272656.	255382.	289930.
2	270491.	248524.	292458.
3	256412.	233235.	279588.
4	276950.	253384.	300516.
5	269348.	245501.	293195.
6	280630.	256459.	304801.
7	275420.	250874.	299966.
8	282663.	257723.	307603.
9	282464.	257133.	307795.
10	276622.	250910.	302335.
11	282886.	256800.	308972.
12	277447.	250995.	303900.
13	279870.	252737.	307002.
14	277205.	249483.	304927.
15	262869.	234682.	291055.
16	283354.	254762.	311946.
17	275776.	246793.	304758.
18	287084.	257712.	316456.
19	281886.	252124.	311647.
20	289130.	258982.	319278.
21	288929.	258399.	319460.
22	283086.	252179.	313994.
23	289350.	258070.	320629.
24	283911.	252264.	315558.

Tabla 6:

Residuales de consumo mensual de energía eléctrica doméstico de la ciudad de Puno

			Límite en 95.0%	Límite en 95.0%
Retraso	Autocorrelación	Error Estd.	Inferior	Superior
1	0.151093	0.0916698	-0.17967	0.17967
2	0.0398019	0.0937392	-0.183726	0.183726
3	-0.135588	0.0938811	-0.184004	0.184004
4	-0.0955918	0.0955125	-0.187202	0.187202
5	0.140766	0.0963132	-0.188771	0.188771
6	-0.0555045	0.0980268	-0.192129	0.192129
7	0.0853251	0.0982905	-0.192646	0.192646
8	-0.000474645	0.098911	-0.193862	0.193862
9	0.0328751	0.098911	-0.193862	0.193862
10	0.0502426	0.0990028	-0.194042	0.194042
11	-0.033516	0.0992168	-0.194462	0.194462
12	-0.100576	0.0993119	-0.194648	0.194648
13	0.00683556	0.100164	-0.196319	0.196319
14	0.0677884	0.100168	-0.196326	0.196326
15	0.048399	0.100553	-0.19708	0.19708
16	0.00837141	0.100748	-0.197464	0.197464
17	0.0339217	0.100754	-0.197475	0.197475
18	0.0755296	0.10085	-0.197663	0.197663
19	-0.00999295	0.101324	-0.198593	0.198593
20	0.0276225	0.101333	-0.198609	0.198609
21	0.0499911	0.101396	-0.198733	0.198733
22	-0.00993507	0.101603	-0.199138	0.199138
23	-0.0768746	0.101611	-0.199154	0.199154
24	-0.0407405	0.102099	-0.20011	0.20011

Tabla 7:

Residuales estimadas de autocorrelación parcial

	Parcial		Límite en 95.0%	Límite en 95.0%
Retraso	Autocorrelación	Error Estd.	Inferior	Superior
1	0.151093	0.0916698	-0.17967	0.17967
2	0.0173694	0.0916698	-0.17967	0.17967
3	-0.147534	0.0916698	-0.17967	0.17967
4	-0.0569117	0.0916698	-0.17967	0.17967
5	0.182839	0.0916698	-0.17967	0.17967
6	-0.124736	0.0916698	-0.17967	0.17967
7	0.0795038	0.0916698	-0.17967	0.17967
8	0.0251848	0.0916698	-0.17967	0.17967
9	0.0245366	0.0916698	-0.17967	0.17967
10	0.0202982	0.0916698	-0.17967	0.17967
11	-0.00303797	0.0916698	-0.17967	0.17967
12	-0.135809	0.0916698	-0.17967	0.17967
13	0.087298	0.0916698	-0.17967	0.17967
14	0.0566108	0.0916698	-0.17967	0.17967
15	-0.0285355	0.0916698	-0.17967	0.17967
16	-0.000635056	0.0916698	-0.17967	0.17967
17	0.100822	0.0916698	-0.17967	0.17967
18	0.0353001	0.0916698	-0.17967	0.17967
19	-0.0373702	0.0916698	-0.17967	0.17967
20	0.0521418	0.0916698	-0.17967	0.17967
21	0.0771118	0.0916698	-0.17967	0.17967
22	-0.0557686	0.0916698	-0.17967	0.17967



	<i>Parcial</i>		<i>Límite en 95.0%</i>	<i>Límite en 95.0%</i>
<i>Retraso</i>	<i>Autocorrelación</i>	<i>Error Estd.</i>	<i>Inferior</i>	<i>Superior</i>
23	-0.102451	0.0916698	-0.17967	0.17967
24	0.0137465	0.0916698	-0.17967	0.17967

MACHINE LEARNING

La intersección de la informática y la neurobiología en la década de 1960 dio lugar a la rama de la inteligencia artificial conocida como aprendizaje automático. Este subcampo de la IA existe desde los años sesenta. Para llevar a cabo sus objetivos, analiza cantidades masivas de datos con el fin de descubrir patrones. No aprende una máquina, sino un algoritmo que ilustra el comportamiento de los datos o saca conclusiones a partir de ellos. Un algoritmo es un conjunto de pasos o pautas que, cuando se siguen, dan como resultado una solución a un problema determinado. Técnica de programación conocida como "aprendizaje automático", cuyo objetivo es enseñar a los ordenadores a aprender a partir de los datos que se les proporcionan. En un sentido puramente técnico, este campo hace posible que los ordenadores aprendan sin necesidad de programación. Por ejemplo, un filtro de correo electrónico o spam es un programa de aprendizaje automático que aprende a detectar spam mostrando ejemplos de spam y comparándolos con otros. **(Cabezón Manchado, 2018).**

El aprendizaje automático nos ayuda en muchas áreas:

- Problemas que tienen solución, pero son muy difíciles de mantener, o dificultades con muchas reglas que dificultan la programación, el aprendizaje automático simplifica el código y logra muchos mejores resultados.
- Problemas complejos sin buenas soluciones. Las nuevas tecnologías pueden encontrar soluciones.



- Problemas donde el entorno cambia. Los sistemas de aprendizaje automático se adaptan a los nuevos datos.

- El aprendizaje automático puede recopilar información sobre problemas muy complejos y grandes cantidades de datos (Cabezón Manchado 2018).

Algunos ejemplos de aprendizaje automático son:

- Predecir la probabilidad de que el usuario retorne dentro de los 30 días.
- Anticipar el tráfico de la ciudad.
- Pronóstico de la demanda de electricidad para el próximo mes.
- Predecir la temperatura media para los próximos 10 días.
- etc.

Existen muchas categorías distintas de algoritmos para el aprendizaje automático, las más comunes de las cuales son el aprendizaje supervisado, el aprendizaje no supervisado, el aprendizaje semisupervisado y el aprendizaje por refuerzo.

APRENDIZAJE SUPERVISADO

El aprendizaje supervisado implica predecir un resultado dadas otras variables y, en este caso, el algoritmo utilizado modela la relación entre la variable predicha y el predictor. Según (Kuhn y Johnson 2013), los modelos de pronóstico involucran la manipulación y partición de valores, así como el desarrollo de modelos matemáticos que producen pronósticos precisos.

Este tipo de aprendizaje es el más utilizado y es el que introducimos en la suite de aprendizaje automático que incluye la solución. Estas soluciones se denominan etiquetas. Las redes neuronales recurrentes son una forma de entrenamiento. Otros ejemplos de este



tipo incluyen: regresión lineal, métodos de k-vecino más cercano, regresión logística, árboles de decisión y bosques aleatorios, entre otros. (Bladehead 2018)

El aprendizaje supervisado se refiere a un tipo específico de modelo de aprendizaje automático en el que el proceso de generación de conocimiento se realiza utilizando un conjunto de ejemplos o datos etiquetados, donde el resultado de la operación se conoce de antemano. Este tipo de modelo aprende de estos resultados y ajusta sus parámetros internos a los nuevos datos que ingresan al sistema.

La predicción usando datos cuantitativos se llama regresión, la clasificación usando datos cualitativos, los ejemplos más comunes de este tipo de aprendizaje son: reconocimiento de voz, traducción de palabras, predicción del precio de la vivienda, predicción de la temperatura, etc.

APRENDIZAJE NO SUPERVISADO

El aprendizaje no supervisado a menudo se realiza como parte del análisis exploratorio de datos, que suele ser más subjetivo y no tiene objetivos analíticos simples, como la predicción de respuestas. Además, puede ser difícil evaluar la calidad de los resultados obtenidos utilizando métodos de aprendizaje no supervisados (Boehmke y Greenwell 2019).

El aprendizaje no supervisado es aquel en el que no hay solución en el conjunto de entrenamiento, es decir, no se etiqueta con lo que el sistema está intentando aprender sin profesor (Cabezón Manchado 2018).

El aprendizaje no supervisado implica conjuntos de datos no etiquetados cuya estructura se desconocía previamente. En este tipo de aprendizaje, el objetivo es explorar una estructura de datos sin etiquetar sin conocimiento previo de la referencia de la variable de salida para extraer información crítica o importante. Hay dos categorías



especiales de este tipo de aprendizaje llamadas agrupación y reducción de dimensionalidad. El agrupamiento es una técnica de investigación utilizada para analizar datos donde la información se organiza en grupos sin un conocimiento previo de la estructura que los compone. La reducción de tamaño con datos muy complejos requiere más capacidad de almacenamiento. El método funciona determinando la correlación entre las características que se encuentran en el conjunto de datos, reduciendo la duplicación de información y reduciendo el tiempo de análisis para una recopilación de información más eficiente y reduciendo el tiempo de análisis para una recopilación de información más eficiente. considerado más valioso.

APRENDIZAJE SEMISUPERVISADO

Es una combinación de aprendizaje previo donde algunos datos están etiquetados (con una respuesta) y otros datos no tienen solución, algunos ejemplos son redes como DEEP BELIEF NETWORKS (DBN). (Bladehead 2018)

El aprendizaje semisupervisado utiliza una pequeña cantidad de datos etiquetados y muchos datos no etiquetados como parte del conjunto de entrenamiento. Este tipo de aprendizaje intenta explorar la información estructural contenida en los datos no etiquetados para producir modelos predictivos que superen a los datos que solo están etiquetados. Usa etiquetas. Los modelos que usan datos etiquetados para construir modelos de clasificación son métodos de autoaprendizaje o aprendizaje orientado a decisiones, donde el primer paso del algoritmo consiste en entrenar un clasificador con una pequeña cantidad de datos etiquetados. Luego, el clasificador se usa para hacer predicciones sobre datos no etiquetados, su predicción más confiable se agrega al conjunto de entrenamiento y, finalmente, el clasificador se vuelve a entrenar usando el nuevo conjunto de entrenamiento (Ibañez y Rubén 2019)



APRENDIZAJE POR REFUERZO

El aprendizaje por refuerzo es un tipo de aprendizaje que se basa en el concepto de la capacidad de un algoritmo para aprender mediante el proceso de ensayo y error. Aunque el resultado ya se conoce de antemano, aún no está claro qué elecciones conducirán al final más satisfactorio. El proceso del algoritmo consiste en asociar y duplicar patrones eficaces hasta perfeccionarlos. El objetivo del aprendizaje por refuerzo es construir modelos que mejoren su rendimiento en función del resultado o recompensa de cada interacción, de la consecuencia de realizar la acción adecuada o del retorno de datos compartidos que alimenten una determinada medida. El modelo maximiza la recompensa del agente utilizando la recompensa como parámetro para ajustar su comportamiento para acciones futuras de modo que las nuevas acciones correspondan al objetivo de acción correcto. En este trabajo de investigación, nos centraremos en el aprendizaje automático de redes neuronales supervisadas.

REDES NEURONALES

Existen neuronas biológicas y neuronas artificiales.

Una neurona es un tipo de célula cerebral que procesa información, recibe impulsos eléctricos a través de sinapsis y envía esa información a otras neuronas. Un cerebro está formado por neuronas, que son células cerebrales que están unidas entre sí.

Hay cuatro partes que componen una neurona, y son las siguientes:

Las dendritas son las partes de una neurona encargadas de recibir los impulsos eléctricos entrantes.

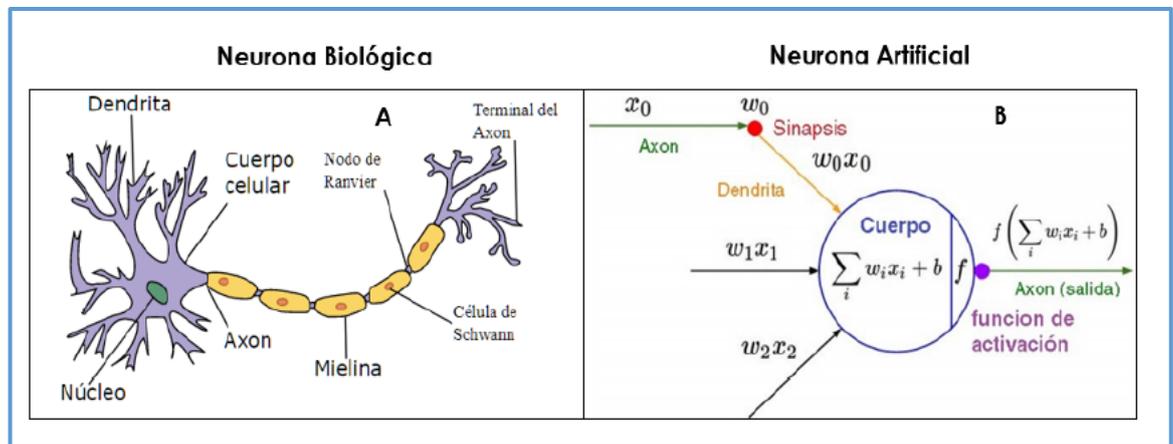
Esta zona se encarga de gestionar la información.

- El axón se encarga de transformar la información entrante en una forma que pueda ser recibida por la neurona que viene detrás.

- La conexión electroquímica entre dos neuronas se denomina sinapsis.

Figura 2:

Neurona biológica y Neurona artificial.



Cada neurona es responsable del aprendizaje a través de su conexión con varias redes de neuronas artificiales. Para ello, recibe señales de otras neuronas, las combina y las envía. Las sinapsis permiten que la información se transfiera entre estas neuronas para que se pueda acceder a ella a través de sus dendritas. Cuando la señal combinada es lo suficientemente fuerte, el nervio descarga un neurotransmisor en el tejido circundante.. Según el tipo de neurotransmisor, las neuronas se excitan cuando se estimulan y se inhiben cuando no se estimulan, produciendo una u otra respuesta según la situación (Cabezón Manchado, 2018). En el diagrama se puede ver una representación de neuronas artificiales.

Dentro de esta neurona sintética, la variable de entrada actúa no sólo como variable explicativa, sino también como variable de entrada. Para que el peso sináptico w_{ij} refleje con precisión la fuerza de la conexión sináptica que existe entre las neuronas, es necesario conectar estas variables con pesos iniciales, que pueden elegirse al azar o permanecer constantes. Neurona presináptica i y neurona postsináptica j . Una neurona es capaz de recibir varias entradas a la vez, y cada una de esas entradas tiene su propio peso



sináptico único., lo que asigna la importancia de la entrada a la función de activación de la neurona. Estos pesos reflejan el estado de la memoria, ya que, si el peso es cercano a cero, dado que no hay conexiones entre neuronas, estos coeficientes se pueden establecer en la red y modificar en respuesta a ejemplos aleatorios de acuerdo con las reglas de entrenamiento (Cabezón Manchado 2018).

REDES NEURONALES ARTIFICIALES O RNAs

Las Redes Neuronales Artificiales (ANNs) están inspiradas en la forma en que funciona el cerebro humano, aunque no logran tales capacidades, son muy útiles en diversas actividades como reconocimiento de voz, tipeo, procesamiento de video, etc. Tiene como objetivo identificar patrones o patrones utilizando métodos algorítmicos tradicionales para resolver diferentes tipos de problemas.

Una red neuronal artificial es un modelo computacional que surgió para lograr una formalización matemática de la estructura del cerebro, imitar la estructura del hardware del sistema nervioso, enfocarse en las funciones del cerebro humano, aprender en base a la experiencia y así obtener conocimiento de ella (Flórez y Fernández 2008).

Una ANN consiste en una colección de neuronas artificiales que forman dispositivos informáticos simples que pueden proporcionar respuestas únicas basadas en vectores de entrada del mundo exterior o estímulos de otras neuronas. Constan de tres tipos de neuronas como se describe (Flórez y Fernández 2008).

Neuronas de entrada: Están ubicadas en la capa de entrada y reciben señales del entorno, sensores u otras partes del sistema.

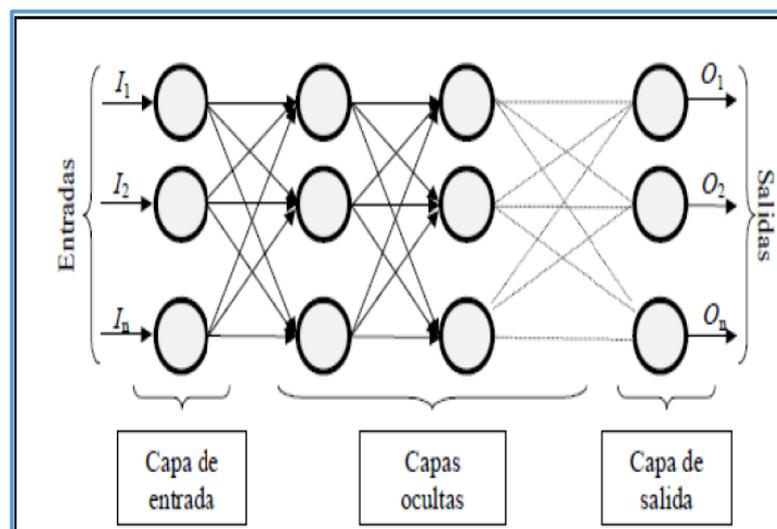
Neuronas de salida: están ubicadas en la capa de salida y envían señales al sistema cuando se completa el procesamiento de la información.

Neuronas ocultas: Se ocultan en la capa oculta, donde reciben la información, la procesan y la transmiten al sistema. Todo esto lo hacen sin interactuar con el mundo exterior.

A continuación, se ilustra en forma de diagrama la estructura esencial de una red neuronal artificial.

Figura 3:

Estructura básica de una red neuronal artificial.



ELEMENTOS BASICOS

Parámetros

Estas etapas son necesarias para entrenar una red neuronal:

Número de capas: En una red neuronal, debe designar el número de capas que desea utilizar. Debe tener al menos capas de entrada y de salida; las capas intermedias son optativas y se utilizan para representar variables de relación no lineal entre redes.

Número de neuronas: También debe especificar el número de neuronas. En la capa de entrada, el número de neuronas viene determinado por el número de entradas o variables independientes, mientras que en la capa de salida, el número de neuronas viene



determinado por el número de salidas o predicciones. El número de neuronas en la capa oculta depende del número de errores producidos por la red; se deben utilizar pruebas para seleccionar las neuronas adecuadas para el tipo de valor que se está procesando. Cabe señalar que no existen reglas específicas con respecto a la cantidad de capas y la cantidad de neuronas para una red neuronal artificial.

Epoch: especifica cuántas veces se repetirán los datos de entrenamiento en la red. Cuanto mayor sea el número de épocas, menor será el error de ajuste.

Batch_size: especifica el tamaño de la submuestra generada para el cálculo del vector de gradiente y la optimización del modelo.

Función de coste: La capacidad de nuestro modelo para proporcionar una estimación precisa de la conexión que existe entre la variable de entrada x y la variable objetivo y está representada por la función de coste. Se espera que el rendimiento de esta función mejore significativamente como resultado de la red. Para variables continuas, algunos ejemplos de medidas de error son el error cuadrático medio, el error absoluto medio y el error porcentual absoluto medio (MSE, MAE y MAPE, respectivamente). Para las variables que son binarias, algunos ejemplos de entropía cruzada son la entropía cruzada binaria y la entropía cruzada de categoría, entre otras.

optimizador: Es un proceso mediante el cual la red se actualiza en función de los datos que ve y de una función de coste; ejemplos de estas funciones son el descenso gradiente y la función de Adam.

TIPOS DE FUNCIONES

Característica de entrada (input function)

El término "entrada global" se refiere a la forma en que varias entradas son procesadas por la neurona como si fueran una sola entrada. Por lo tanto, lo que se requiere

es combinar las entradas (x_{i1}, x_{i2}, \dots) , dentro de la entrada global, x_{ij} indica que la variable explicativa x_j entra a la neurona i . Esto se logra a través de la función de la entrada, la cual se calcula a partir del vector de entrada. La función de entrada puede describirse como sigue:

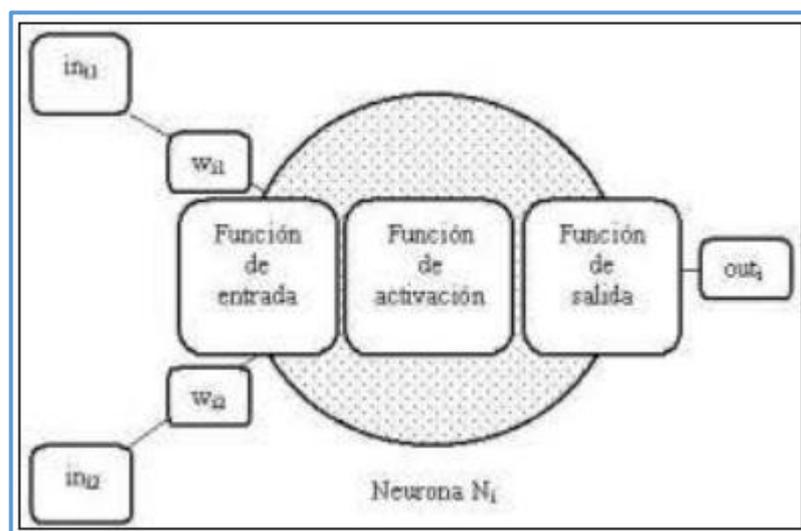
$$input_i = (x_{i1} w_{i1}) * (x_{i2} w_{i2}) * \dots * (x_{in} w_{in})$$

Donde: $input_i$ es la entrada a la neurona i , * representa al operador apropiado que puede ser la suma, el producto, el máximo, etc., n es el número de entradas a la neurona N_i y w_i es el peso.

En general, los pesos ilimitados cambian la magnitud de la influencia de los valores de entrada, es decir, permiten que los valores de entrada más grandes tengan solo un efecto pequeño si los valores de entrada son lo suficientemente pequeños. (Matich, 2001).

Figura 4:

Neurona con dos entradas y una salida.





Función de activación

Una neurona artificial tiene un estado de activación que puede ser activa (excitada) o inactiva (no excitada), la función de activación determina este estado transformando la entrada global en un valor que usualmente este dado en un intervalo de (0 o 1) o de (-1 a 1), cuando una neurona está totalmente inactiva toma valores de (0 o -1) y cuando esta activa toma el valor de (1).

Esta función de activación f_i procesa el estado de actividad de una neurona. Define el estado de activación actual $a_i(t)$ de la neurona en base al potencial resultante h_i y al estado de activación anterior a la neurona $a_i(t - 1)$, el estado de activación de una neurona i para un determinado instante de tiempo t puede ser expresado de la siguiente manera:

$$a_i(t) = f_i(a_i(t - 1), h_i(t))$$

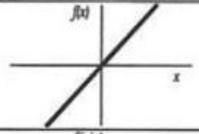
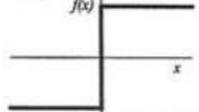
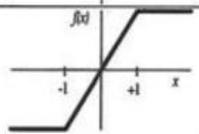
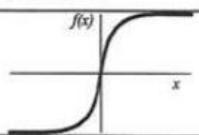
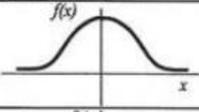
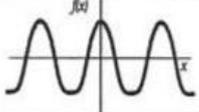
En la mayoría de los casos se suele ignorar el estado de activación anterior de la neurona quedando de la siguiente manera:

$$a_i(t) = f_i(h_i(t))$$

Las funciones de activación más utilizadas son:

Figura 5:

Funciones de activación habituales.

	Función	Rango	Gráfica
Identidad	$y = x$	$[-\infty, +\infty]$	
Escalón	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, +1\}$ $\{0, +1\}$	
Lineal a tramos	$y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } -l \leq x \leq l \\ +1, & \text{si } x > l \end{cases}$	$[-1, +1]$	
Sigmoidea	$y = \frac{1}{1+e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	
Gaussiana	$y = Ae^{-Bx^2}$	$[0, +1]$	
Sinusoidal	$y = A \text{sen}(\omega x + \varphi)$	$[-1, +1]$	

Función de salida (output function)

El componente final de una neurona es la función de salida; los datos derivados de esta función son la salida de la neurona i (*out*); por consiguiente, la función de salida determina qué datos se transmiten a las neuronas conectadas. Si la función de activación está por debajo de un determinado umbral, no se transmite ninguna salida a la neurona siguiente. Normalmente, no se permite ningún dato como salida pasa una neurona, por lo que los valores de salida están en el rango $[0,1]$ o $[-1,1]$, o pueden ser binarios $0,1$ o $-1,1$.

Tenemos dos de las funciones de salida más comunes son:

Identidad, este es de tipo de función más sencillo, tal que la salida es la misma que la entrada.



Binaria, $\begin{cases} 1 & \text{si } a_i(t) \geq \varepsilon_i \\ 0 & \text{de lo contrario} \end{cases}$; donde ε_i es el umbral de aceptación.

Aprendizaje de una red neuronal

Las redes neuronales manejan dos tipos de información, la primera es información volátil que se refiere a los datos con los que están trabajando y que cambia a medida que cambia la dinámica de la red. Esta información se almacena en el estado dinámico de la neurona. El segundo tipo es información volátil que se retiene para recordar patrones aprendidos y se almacena en pesos sinápticos.

El proceso de aprendizaje o entrenamiento de una red neuronal implica encontrar relaciones entre los pesos para realizar la tarea para la que está diseñada la red, como la predicción o la clasificación.

El aprendizaje se basa en entrenar la red con patrones, estos patrones a menudo se denominan muestras o patrones de entrenamiento, el proceso estocástico del algoritmo es que la red ejecuta iterativamente los patrones cambiando los pesos sinápticos hasta que convergen en un conjunto óptimo de pesos, representando estos patrones lo suficientemente bien, es decir, los pesos sinápticos son adecuados para responder correctamente al conjunto de patrones de entrenamiento que les mostramos. (Cabezón Manchado 2018).

El aprendizaje de ARN requiere técnicas de minimización de errores u optimización de la función de recompensa para cambiar los valores de los pesos sinápticos de acuerdo con los datos de entrada disponibles, optimizando así la respuesta de la carne de res a nuestro resultado deseado, el objetivo del proceso de entrenamiento es encontrar Pesos seleccionados para minimizar la función de error.

Las etapas del proceso de aprendizaje son:



- Inicialización de la red: comienza el aprendizaje, se generan pesos aleatorios para todas las conexiones en $n = 0$. Usando estos pesos, se puede calcular el primer error de predicción o clasificación.

- Hacia adelante: La información se envía unidireccionalmente de entrada a salida utilizando funciones de activación y pesos. Para cada tiempo n , se genera un nuevo modelo de entrada y los pesos se refinan para mejorar la tasa de error en el paso anterior.

- Distribución de errores: El algoritmo se detiene cuando el error alcanzado es inferior a un límite preestablecido, entonces el error simplemente no disminuye más o se alcanza el número de iteraciones, de lo contrario pasa al siguiente paso.

- Propagación de errores: El error obtenido en la capa de salida se propaga hacia atrás para cambiar los pesos. - Ajuste: Los pesos se ajustan mediante el método de Adam o el método de gradiente descendente para reducir el error. (Color de color Kabuzzón2018)

Tipos de redes neuronales

Las redes neuronales artificiales se pueden clasificar por topología, es decir, según el número de capas que pueden ser de entrada, salida u ocultas, y según la dirección de las conexiones neuronales, las redes se clasifican en:

Redes de una sola capa: Son redes de una sola capa. Las conexiones se realizan entre las neuronas en una capa para formar una conexión, y una red de una sola capa tiene neuronas que pueden conectarse entre sí, llamadas bucles automáticos.

Redes multicapa: estas redes tienen múltiples capas de neuronas. Además, se pueden clasificar por el tipo de conexión de las capas:



- Feedforward: también conocidas como estáticas, estas redes tienen conexiones de capa unidireccionales (directas) desde la entrada hasta la salida, incluidos perceptrones, redes adaptativas, etc.

Feedback: estas redes, también conocidas como dinámicas, tienen conexiones hacia adelante y hacia atrás entre las capas para que la información pueda retroalimentarse a las capas anteriores de la dinámica de la red. (Barhead 2018)

Redes neuronales recurrentes

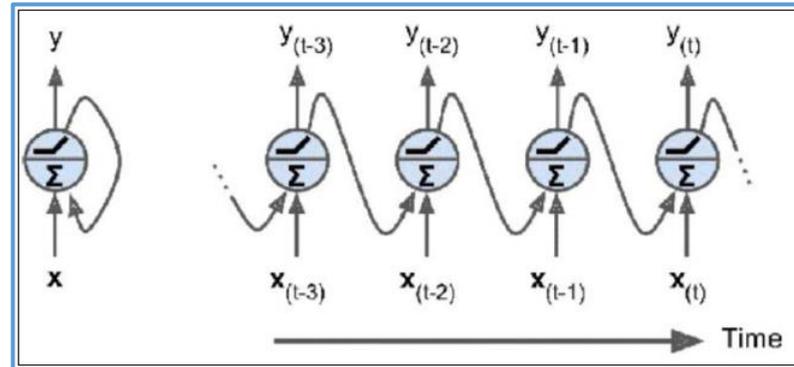
Las redes neuronales recurrentes (RNN) son redes que contienen bucles que almacenan información como una especie de memoria. Estas redes carecen de una estructura escalonada específica, lo que permite que las neuronas se enlacen arbitrariamente entre sí o consigo mismas. Las redes neuronales recurrentes se utilizan en inteligencia artificial y aprendizaje automático. Además del análisis de texto, audio y vídeo, son útiles para predecir el orden de los acontecimientos en una secuencia.

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.

Estas redes funcionan de la misma manera que cuando una persona lee un texto, conoce y comprende las palabras que ve porque ha almacenado el significado de las palabras en la memoria, al igual que las RNN almacenan información sobre eventos pasados.

Figura 6:

Neurona recurrente (izquierda), desarrollo en el tiempo (derecha).



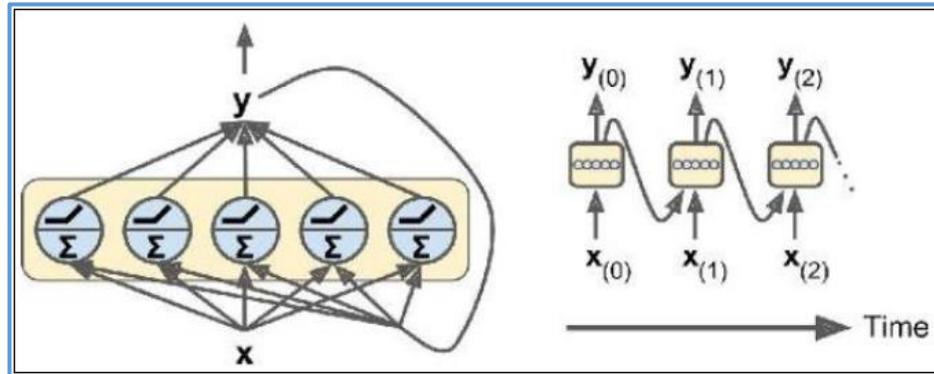
En las redes neuronales recurrentes (RNN), las salidas de las neuronas están correlacionadas con los datos variables actuales que se introducen en la red, así como con los datos de entrada anteriores, los datos de salida anteriores, etc.

En la figura N°13, se puede observar que cada paso temporal t , esta neurona recibe las x_t como la salida del paso anterior, $y_{(t-1)}$, si representamos este proceso en el eje x como si fuera el tiempo, obtendríamos la parte de la figura N°13.

Con esto se puede formar fácilmente una capa con neuronas recurrentes, en cada tiempo t , todas las neuronas reciben el vector de entrada $x_{(t)}$ y el vector de salida del paso anterior $y_{(t-1)}$, como se muestra a continuación:

Figura 7:

Capa de neuronas recurrentes (izquierda), desarrollo en el tiempo (derecha).



Cada neurona recurrente tiene dos conjuntos de pesos, uno de los inputs $x_{(t)}$ y otro para las salidas de paso temporal anterior $y_{(t-1)}$. Llamaremos a estos pesos w_x y w_y , la salida $y_{(t)}$ de una neurona recurrente se puede calcular de la siguiente manera:

$$y_{(t)} = \phi (x_{(t)}^T * w_x + y_{(t-1)}^T * w_y + b)$$

Donde b es el sesgo (bias) y ϕ es la función de activación. Podemos resumir la expresión mediante la expresión anterior mediante el álgebra lineal de la siguiente manera:

$$Y_{(t)} = \phi (X_{(t)} * W_x + Y_{(t-1)} * W_y + b) = \phi ([X_{(t)} Y_{(t-1)}] * W + b)$$

Siendo $W = [W_x \ W_y]^T$ y donde:

$Y_{(t)}$: es una matriz $m \times n_{inputs}$ que contiene los outputs del marco temporal t para cada *batch* (conjunto de datos que entra a la neurona) en este caso $m \times n_{inputs}$ es el número de neuronas de la capa.

$X_{(t)}$: es una matriz $m \times n_{inputs}$ que contiene las variables inputs de las m observaciones y n_{inputs} es el número de variables explicativas de entrada.

W_x : es una matriz $n_{inputs} \times n_{neuronas}$ conteniendo los pesos de las conexiones para las entradas del paso temporal actual.

W_y : es una matriz $n_{neuronas} \times n_{neuronas}$ conteniendo los pesos de las conexiones para las salidas del paso temporal actual.

Las matrices de pesos W_x y W_y se pueden juntar en la matriz W que será $(n_{inputs} \times n_{neuronas}) \times n_{neuronas}$

b es un vector de 1's de tamaño $n_{neuronas}$ que permite añadir un término constante (*bias*) al modelo.

Debemos darnos cuenta de que $Y_{(t)}$ es una función de $X_{(t)}$ y $Y_{(t-1)}$, quien a su vez es función de $X_{(t-1)}$ y de $Y_{(t-2)}$ quién es función de $X_{(t-2)}$ y de $Y_{(t-a)}$, etc. Esto hace que $Y_{(t)}$ sea función de todas las variables de entrada desde el tiempo inicial $t = 0$ (esto es, de $X_{(0)}, X_{(1)}, \dots, X_{(t)}$), en $t = 0$, no hay outputs anteriores luego se asumen como ceros (Cabezón Manchado 2018).

Red de Elman

La red de Elman es la más simple. Su estructura es la base de otras redes neuronales recurrentes. Su estructura consta de capa de entrada, capa oculta y capa de salida. Una capa tiene una unidad de contexto, también llamada unidad de retardo. El número de neuronas en esta capa debe ser igual al número de neuronas en la capa oculta.

La arquitectura de la red tiene retroalimentación que permite la volatilidad, lo que asegura que la red tenga memoria y que el tipo de memoria sea de corta duración. El proceso de memoria se realiza a través de las unidades de retardo proporcionadas por las neuronas de la capa oculta, el peso de conexión entre la capa oculta y la unidad de retardo es fijo e igual a 1, lo que permite obtener siempre una copia de la salida de la capa oculta.

capa con el valor de la salida de la capa oculta. paso previo (Cabezón Manchado 2018).

La ecuación matemática que define el proceso es:

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$

$$y_t = \sigma_y(W_y h_t + b_y)$$

Donde x_t , es el vector con las variables de entrada, h_t es el vector de la capa oculta, y_t es el vector de salida, $W = [W_x \ W_y]^T$, y U_h son las matrices de pesos y $b = [b_h \ b_y]^T$, el bias, σ_h y σ_y son las funciones de activación de las respectivas capas (Cabezón Manchado 2018).

APLICACIÓN REDES NEURONALES RECURRENTE

El tipo de escala afecta el procesamiento de las redes neuronales profundas, especialmente cuando se utilizan funciones de activación de tangentes sigmoideas o hiperbólicas. (Cabezón Manchado 2018).

Otra normalización es escalar las funciones entre un valor mínimo y máximo dado (en la práctica se utilizan los rangos [-1,1] y [0,1]) para que cada variable se escale al mayor valor absoluto del tamaño de la unidad.

La fórmula que se utiliza para escalar una variable aleatoria y_i es:

$$z_i = \frac{y_i - \min(Y)}{\max(Y) - \min(Y)}$$

Donde $\min(Y)$ y $\max(Y)$ son valores mínimos y máximos del vector Y y z_i está escalado entre 0 y 1.

La siguiente formula se usa para transformar los datos estandarizados a la escala original:

$$y_i = z_i * (\max(Y) - \min(Y)) + \min(Y)$$



Una vez que los valores están listos para usar, comenzamos a construir la red neuronal recurrente.

En este caso, se implementará utilizando redes celulares, y cada red tiene una infinidad de parámetros con los que puedes experimentar hasta obtener mejores resultados, ya que puedes usar funciones de activación, funciones de reducción de pérdidas, como la cantidad de capas. el número de neuronas en cada capa, etc. Este trabajo de investigación muestra en detalle la red que da mejores resultados. Antes de crear la red, se describirá en detalle el funcionamiento de la técnica Dropout.

Dropout es un método para eliminar o inhabilitar neuronas en una red neuronal con cierta probabilidad. Este parámetro tiene un valor entre 0 y 1. Si el valor es cercano a 0, se activarán menos neuronas. Un valor cercano a 1 deshabilitará múltiples variables.

Este proceso nos ayuda a evitar el sobre entrenamiento, ya que las capas totalmente conectadas capturan la mayoría de los parámetros y, por lo tanto, las neuronas desarrollan interdependencias entre ellas, lo que limita el poder individual de cada neurona y provoca el sobre entrenamiento (Cabezón Manchado 2018).

En este estudio, a la deserción se le asignó un valor de 0,2 porque el estudio concluyó que se desempeñó mejor, y la red neuronal se probó sin deserción y se desempeñó peor.

Es necesario segmentar la serie temporal en muestras que incluyan componentes de entrada y salida si la serie temporal se va a utilizar para la previsión multinivel. Cada uno de los componentes de entrada y salida estará formado por muchos pasos temporales, aunque el número de pasos de cada componente puede o no ser el mismo. Tomemos, por ejemplo, el caso de una serie temporal con una sola dimensión.

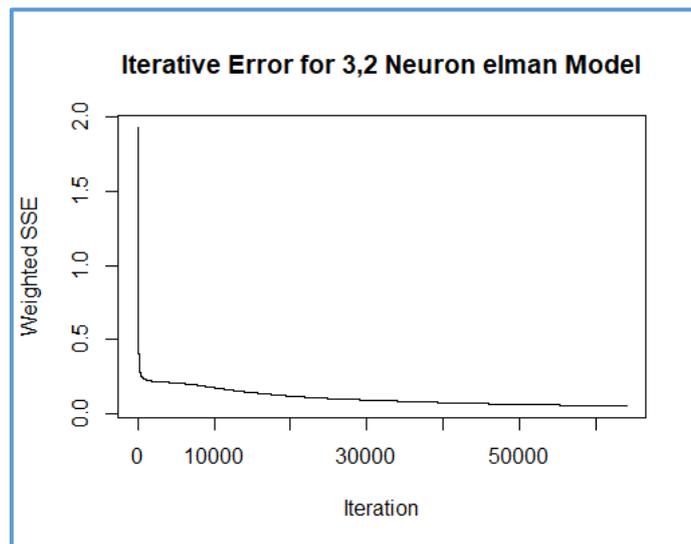
$$Y = \{1,2,3,4,5,6,7,8,9,10\}$$

inputs outputs

<i>muestra 1</i>	→	$\begin{bmatrix} 1, 2, 3 \end{bmatrix}$	$\begin{bmatrix} 4, 5 \end{bmatrix}$
<i>muestra 2</i>	→	$\begin{bmatrix} 2, 3, 4 \end{bmatrix}$	$\begin{bmatrix} 5, 6 \end{bmatrix}$
<i>muestra 3</i>	→	$\begin{bmatrix} 3, 4, 5 \end{bmatrix}$	$\begin{bmatrix} 6, 7 \end{bmatrix}$
<i>muestra 4</i>	→	$\begin{bmatrix} 4, 5, 6 \end{bmatrix}$	$\begin{bmatrix} 7, 8 \end{bmatrix}$
<i>muestra 5</i>	→	$\begin{bmatrix} 5, 6, 7 \end{bmatrix}$	$\begin{bmatrix} 8, 9 \end{bmatrix}$
<i>muestra 6</i>	→	$\begin{bmatrix} 6, 7, 8 \end{bmatrix}$	$\begin{bmatrix} 9, 10 \end{bmatrix}$

Figura 8:

Comportamiento del error de la red de Elman.



En la figura N°26, se representa el comportamiento del error en función al número de iteraciones aproximadamente el error de la iteración uno y dos es grande y a partir de la iteración tres el error se estabiliza o se mantiene.

Villareal(2006), se desarrollaron varias RNA para estimar el desempeño de predicción, seleccionar el método de pronóstico más competitivo dadas dos características de la serie de tiempo bajo análisis, el número de datos históricos disponibles (t) y el orden máximo de un polinomio (n) ajustado a la serie de tiempo con un nivel de aproximación de mas de 80%.

El presente trabajo aplica este método a series de tiempo reales.

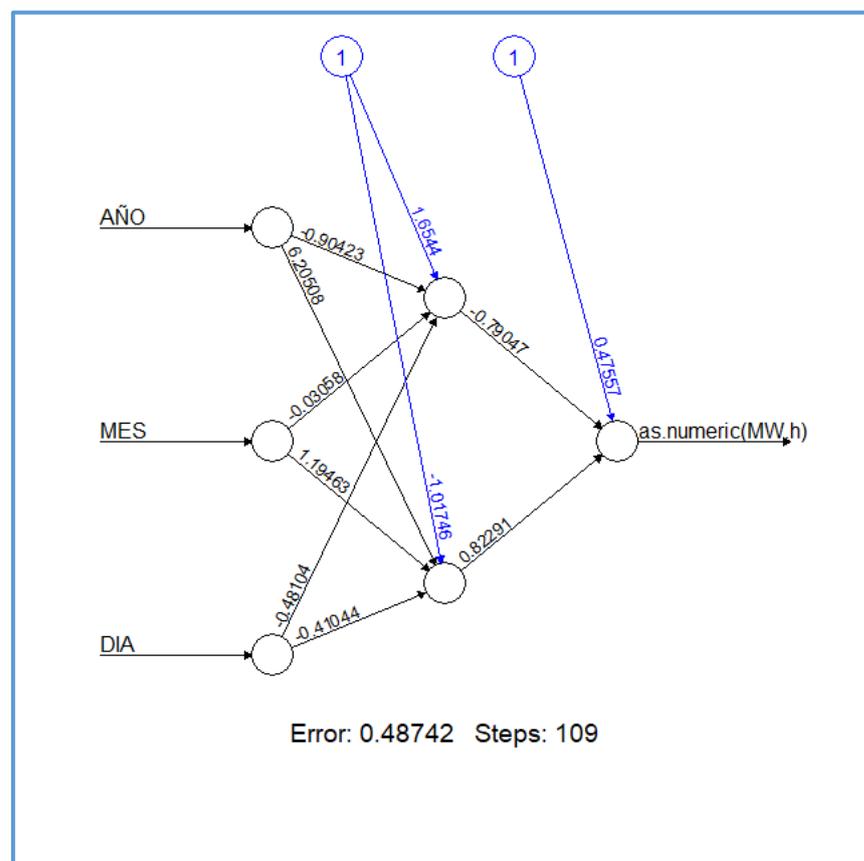
Escalar los datos al intervalo $[-1,1]$. Los valores de la serie de tiempo se normalizan para que caigan en un rango $[-1,1]$ con el objetivo de eliminar efectos de dimensionalidad.

Caracterizar la serie de tiempo, se determinan dos parámetros que caracterizan la serie; el número de periodos de la serie de tiempo (t) y el grado del primer polinomio (n) que se ajuste a los datos serie con un coeficiente de determinación, $R^2 > 80\%$.

Someter a la red neuronal artificial. Se utiliza una red neuronal artificial previamente entrenada, donde se usan como entradas t y n como salida la predicción del error cuadrado medio para cada uno de los métodos de pronóstico.

Figura 26:

Redes Neuronales valores de entrada y salida.

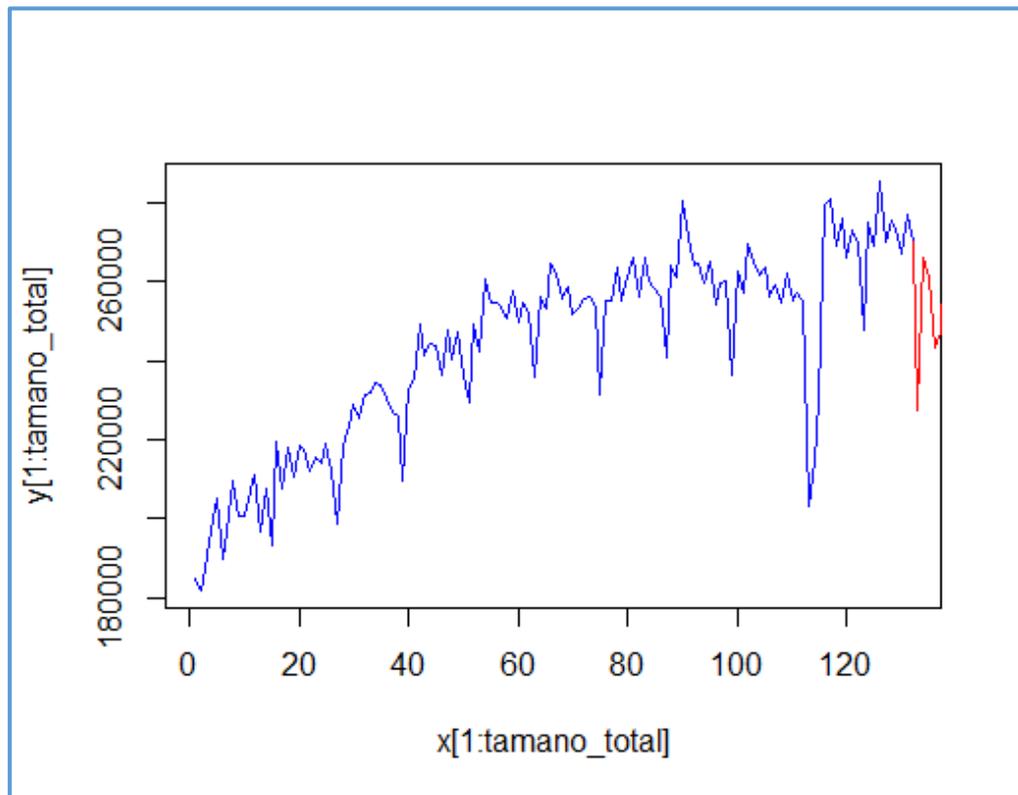




En la figura N°26, se puede observar que tenemos tres valores de entrada y con un valor de salida (MW-h/mes), este sería el valor de cada registro a predecir el cual tendrá al momento de salir la misma dimensión de los valores de entrada, todo lo que está en medio no será así las funciones que se observa (capas ocultas) no tendrán la misma dimensión porque, es donde entra la función de activación, las líneas que se observa son las que se encargan en la conexión o la dirección entre las variables de entrada y las neuronas (capas del medio), también podemos observar los análogos (bayas 1, bayas 1) que están conectadas a las redes neuronales y la capa de salida de color azul, que vendría hacer la rectificación o valor constante el cual mejora la capacidad de predicción y se crea como un apoyo a la capacidad de pronóstico de cada uno de los valores de entrada; se puede decir también que la variable entrada año, vayas 01 y capa oculta son más resaltantes porque lograron pasar las ecuaciones más fuertes en las redes neuronales al hacer la predicción.

Figura 27:

Predicciones realizadas RNR.



En la figura N°27, Podemos observar que los valores tratan por replicar la tendencia que sube y baja no logra en ningún momento apegarse a los datos reales por que estos datos son de predicción y los reales son los totales, podemos concluir que la red neuronal trata apegarse a lo más posible a los valores reales no lo logra apegarse del todo es por eso que el error absoluto y porcentual son considerablemente altos.



```
#SERIES DE TIEMPO METODO #  
ELECTRO_PUNO_1  
Arimar.ts=ts(ELECTRO_PUNO_1, star=c(2011,1),end=c(2021,12),frequency=12)  
Arimar.ts  
plot(Arimar.ts,xlab="AÑO",col="brown2", type="o",pch=20, ylab="Consumo de Energia")  
grid()  
#probando dicky fuller#  
adf.test(Arimar.ts)  
plot(Arimar.ts)  
acf(Arimar.ts)  
pacf(Arimar.ts)  
#primera dieferencia#  
DY<-diff(Arimar.ts,lag = 1)  
adf.test(DY)  
plot(DY)  
acf(DY)  
pacf(DY)  
#MODELO ARIMA#  
modelo<-auto.arima(DY,trace = T,stepwise = F,approximation = F,max.d = 0)  
print(modelo)
```



```
#para pronosticar#  
modelo_arma<-Arima(Arimar.ts, order = c(0,0,3), seasonal = list(order = c(2,0,0),period=12))  
modelo_arma  
summary(modelo_arma)  
coeftest(modelo_arma)  
accuracy(modelo)  
lines(fitted(modelo),col="red")  
tsdiag(modelo)  
Box.test(residuals(modelo),type = "Ljung-Box")  
error=residuals(modelo)  
plot(error)  
ggtsdiag(modelo)  
summary(modelo)  
autoplot(acf(modelo$residuals,plot = FALSE))  
autoplot(pacf(modelo$residuals,plot = FALSE))  
pronostico<-forecast(modelo_arma,level = c(95),h=12)  
plot(pronostico)  
pronostico  
#APLICANDO REDES NEURONALES#  
library(RSNNS)  
#viendo caracteres de datos#  
library(neuralnet)  
library(NeuralNetTools)
```



```
plot(ELECTRO_PUNO_1)

tsData = ts(ELECTRO_PUNO_1, start = c(2011,1), end = c(2021,12), frequency = 12 )

print(tsData)

plot(tsData)

stsData <- as.ts(tsData,F)

stsDataN <- (stsData - min(stsData) )/(max(stsData) - min(stsData))

plot(stsDataN)

set.seed(1993)

tamano_total <- length(stsDataN)

tamano_train <- round(tamano_total*0.75, digits = 0)

train <- 0:(tamano_train-1)

test <- (tamano_train):tamano_total

test

set.seed(1993)

fit <- elman(inputs[train], outputs[train], size=c(7,3),learnFuncParams=c(0.1),maxit=64000)

plotIterativeError(fit, main = "Iterative Error for 3,2 Neuron elman Model")

y<-as.vector(outputs[-test])

plot(y,type = "l")

pred <- predict(fit, inputs[-test])

lines(pred,col = "red")

predictions <- predict(fit, inputs[-train])

mod5 <- predictions*(max(stsData) - min(stsData) ) + min(stsData)
```



```
mod5
x <- 1:(tamano_total+length(mod5))
y <- c(as.vector(tsData),mod5)
plot(x[1:tamano_total], y[1:tamano_total], col = "blue", type="l")
lines( x[(tamano_total):length(x)], y[(tamano_total):length(x)], ,col="red")
n <- nrow(energia)
muestra <- sample(n,n*0.7)
train <- energia[muestra, ]
test <- energia[-muestra,]
# NORMALIZACION DE VARIABLES
maxs <- apply(train, 2, max)
mins <- apply(train, 2, min)
energia_nrm <- as.data.frame(scale(energia, center = mins, scale = maxs - mins ))
train_nrm <- energia_nrm [muestra, ]
test_nrm <- energia_nrm [-muestra,]

#FORMULA#

nms <- names(train_nrm)
frml <- as.formula(paste("as.numeric(MW.h) ~ ",paste(nms[!nms %in% "MW.h"], collapse = "
+ ")))
#MODELO#
modelo.nn <- neuralnet(frml,
```



```
data      = train_nrm,  
hidden    = c(2),  
threshold = 0.05,  
algorithm  = "rprop+",  
)  
modelo.nn  
plot(modelo.nn)  
summary(modelo.nn)  
pr.nn <- compute(modelo.nn,within(test_nrm,rm(MW.h)))  
pr.nn  
# se transoforma el valor escalar al valor nominal original  
camp.predict      <-      pr.nn$net.result*(max(energia$MW.h)-  
min(energia$MW.h))+min(energia$MW.h)  
camp.real          <-      (test_nrm$MW.h)*(max(energia$MW.h)-  
min(energia$MW.h))+min(energia$MW.h)  
camp.predict  
camp.real  
# S. D. E. C.  
(se.nn <- sum((camp.real - camp.predict)^2)/nrow(test_nrm))  
# Errores  
qplot(x=camp.real, y=camp.predict, geom_point("lm"),  
      main=paste("Real Vs Prediccion. Suma de Error Cuadratico=", round(se.nn,2)))
```



DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo **GERBERT BEKENBAVER QUISPE QUISPE**,
identificado con DNI: **73977671**, en mi condición de egresado de:

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

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

**“MODELO UNIVARIANTE DE CONSUMO MENSUAL DE ENERGÍA ELÉCTRICA
DOMESTICA DE LA CIUDAD DE PUNO – ELECTRO PUNO, PERIODO 2011 - 2021”**

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 Agosto del 2023



FIRMA (obligatoria)



Huella



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

Por el presente documento, Yo **GERBERT BEKENBAVER QUISPE QUISPE**,
identificado con DNI **73977671**, en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado

INGENIERIA ESTADÍSTICA E INFORMÁTICA,

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

**“MODELO UNIVARIANTE DE CONSUMO MENSUAL DE ENERGÍA ELÉCTRICA DOMÉSTICA
DE LA CIUDAD DE PUNO – ELECTRO PUNO, PERIODO 2011 - 2021”**

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 Agosto del 2023



FIRMA (obligatoria)



Huella