

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO

FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICAS

ESCUELA PROFESIONAL DE MATEMÁTICA



TESIS

**METODOLOGÍA BOX JENKINS Y RED NEURONAL
ARTIFICIAL EN LA PREDICCIÓN DE LA PRODUCCIÓN DE
AZÚCAR EN EL PERÚ, 1994 - 2023**

PRESENTADO POR:

Br. ABDIL MICHELL RAMOS HUAYCHO

**PARA OPTAR AL TÍTULO PROFESIONAL
DE LICENCIADO EN MATEMÁTICA
MENCIÓN ESTADÍSTICA**

ASESOR:

Dr. ALFREDO VALENCIA TOLEDO

CUSCO – PERÚ

2025



Universidad Nacional de San Antonio Abad del Cusco

INFORME DE SIMILITUD

(Aprobado por Resolución Nro.CU-321-2025-UNSAAC)

El que suscribe, el Asesor ALFREDO VALENCIA TOLEDO
..... quien aplica el software de detección de similitud al
trabajo de investigación/tesis titulada: METODOLOGÍA BOX JENKINS Y RED
NEURONAL ARTIFICIAL EN LA PREDICCIÓN DE LA PRODUCCIÓN DE
AZÚCAR EN EL PERÚ, 1994-2023

Presentado por: ABDIL MICHEL RAMOS HUAYCHO DNI N° 75981503
presentado por: DNI N°:
Para optar el título Profesional/Grado Académico de LICENCIADO EN MATEMÁTICA
MENTIÓN ESTADÍSTICA

Informo que el trabajo de investigación ha sido sometido a revisión por 2 veces, mediante el
Software de Similitud, conforme al Art. 6° del **Reglamento para Uso del Sistema Detección de**
Similitud en la UNSAAC y de la evaluación de originalidad se tiene un porcentaje de 4 %.

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No sobrepasa el porcentaje aceptado de similitud.	X
Del 11 al 30 %	Devolver al usuario para las subsanaciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, conforme al reglamento, quien a su vez eleva el informe al Vicerrectorado de Investigación para que tome las acciones correspondientes; Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	

Por tanto, en mi condición de Asesor, firmo el presente informe en señal de conformidad y adjunto
las primeras páginas del reporte del Sistema de Detección de Similitud.

Cusco, 18 de noviembre de 2025

Alfredo Valencia Toledo
Firma
Post firma Alfredo Valencia Toledo
Nro. de DNI 43162177
ORCID del Asesor 000-0001-6505-9634

Se adjunta:

- Reporte generado por el Sistema Antiplagio.
- Enlace del Reporte Generado por el Sistema de Detección de Similitud: oid: 27259:529566799

ABDIL MICHELL RAMOS HUAYCHO

METODOLOGÍA_BOX_JENKINS_Y_RED_NEURONAL_ARTIFICIAL_ - 2023.pdf

 Universidad Nacional San Antonio Abad del Cusco

Detalles del documento

Identificador de la entrega

trn:oid:::27259:529566749

Fecha de entrega

18 nov 2025, 1:05 p.m. GMT-5

Fecha de descarga

18 nov 2025, 1:32 p.m. GMT-5

Nombre del archivo

METODOLOGÍA_BOX_JENKINS_Y_RED_NEURONAL_ARTIFICIAL_EN_LA_PREDICCIÓN_DE_LA_PRODU....pdf

Tamaño del archivo

4.1 MB

120 páginas

22.836 palabras

121.788 caracteres

4% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...




Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 15 palabras)

Exclusiones

- N.º de coincidencias excluidas

Fuentes principales

- 3%  Fuentes de Internet
- 0%  Publicaciones
- 3%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

DEDICATORIA

A Dios todo poderoso que guía mi camino,
me da días felices, oportunidades para aprender,
sabiduría y salud.

A mis padres por sus sabios consejos y el ánimo para seguir adelante, también a mis
hermanos por estar conmigo mientras realizaba el trabajo.

AGRADECIMIENTO

Primeramente, agradezco a Dios todo poderoso por guiar mis pasos en todo el transcurso de mi vida, por las oportunidades que me dio de poder formarme profesionalmente y personalmente, manteniéndome con salud y permitiendo que siga aprendiendo dándome sabiduría.

A mis padres, por su amor incondicional, su apoyo en cada paso de mi vida y por enseñarme el valor del esfuerzo, los consejos que me dieron fueron una pieza fundamental en el transcurso de mi carrera y mi vida personal.

A mis hermanos por sus buenos consejos, sus apoyos emocionalmente y las alegrías compartidas en momentos difíciles.

Al Dr. Alfredo Valencia Toledo, docente del Departamento Académico de Matemática y Estadística de la Universidad Nacional de San Antonio Abad del Cusco, por la paciencia y el tiempo dedicado en la asesoría del presente trabajo.

De la misma forma agradezco a todos los docentes de la escuela profesional de Matemática y Estadística por las enseñanzas académicas que recibí y la amistad como los consejos que me impartieron.

PRESENTACIÓN

SEÑOR DECANO DE LA FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y
MATEMÁTICAS

SEÑOR DIRECTOR DE LA ESCUELA PROFESIONAL DE MATEMÁTICA

SEÑORES DOCENTES MIEMBROS DEL JURADO

De acuerdo con el reglamento de Grados y Títulos establecidos por la Escuela Profesional de Matemática Mención Estadística, presento a vuestra consideración el presente trabajo de tesis titulado METODOLOGÍA BOX JENKINS Y RED NEURONAL ARTIFICIAL EN LA PREDICCIÓN DE LA PRODUCCIÓN DE AZÚCAR EN EL PERÚ, 1994 – 2023, conducente al título profesional de Licenciado en Matemática Mención Estadística.

En la presente investigación se propone encontrar el modelo adecuado para la predicción de la producción de azúcar, utilizando la metodología de Box Jenkins y las redes neuronales artificiales, lo mismo que tendrá utilidad para las entidades dedicadas al sector del azúcar en el país, para lo cual se utilizó el software R-Project versión 4.3.1.

Por último, la presente investigación tiene como finalidad comparar ambas metodologías para ver el mejor modelo de predicción.

Atentamente

Br. Abdil Michell Ramos Huaycho

RESUMEN

La presente investigación tiene como objetivo analizar y comparar la capacidad predictiva de la metodología Box Jenkins y de una red neuronal artificial (RNA) en la estimación de la producción de azúcar en el Perú, en el periodo 1994 - 2023. Este estudio responde a la necesidad de contar con modelos confiables que permitan anticipar el comportamiento de un sector agrícola - industrial estratégico para la economía nacional.

Se implementó un modelo de RNA de tipo perceptrón multicapa, con 13 neuronas en la capa oculta y una capa de salida, obteniendo un Error Porcentual Absoluto Medio (MAPE) de 12.064%. Dicho resultado evidencia la capacidad de las RNA para modelar relaciones no lineales; aunque la magnitud del error obtenido limita su precisión. En contraste, el modelo SARIMA (3, 0, 1) (0, 1, 1)₁₂ perteneciente a la metodología Box Jenkins, alcanzó un MAPE de 3.879%, demostrando un mejor desempeño y un ajuste más adecuado a la estacionalidad y variabilidad de la serie.

El procesamiento de los datos realizado en el software R-Project garantiza la rigurosidad en la estimación de parámetros y en la comparación de resultados. Los hallazgos concluyen que, si bien las redes neuronales constituyen una alternativa moderna y flexible en la predicción de series temporales, en este caso el modelo Box Jenkins ofrece mayor precisión y confiabilidad. Así, los resultados obtenidos representan un insumo relevante para la planificación agrícola, la gestión de recursos y formulación de políticas orientadas al fortalecimiento de la industria azucarera en el país.

Palabras claves: Metodología Box Jenkins, perceptrón multicapa, red neuronal artificial, SARIMA.

ABSTRACT

The present research aims to analyze and compare the predictive performance of the Box–Jenkins methodology and an artificial neural network (ANN) in estimating sugar production in Peru for the period 1994–2023. This study responds to the need for reliable models capable of anticipating the behavior of an agricultural–industrial sector that is strategic for the national economy.

An ANN model of the multilayer perceptron type was implemented, with 13 neurons in the hidden layer and one output layer, obtaining a Mean Absolute Percentage Error (MAPE) of 12.064%. This result demonstrates the ANN's ability to capture nonlinear relationships, although the magnitude of the error limits its predictive accuracy. In contrast, the SARIMA $(3,0,1)(0,1,1)_{12}$ model from the Box–Jenkins methodology achieved a MAPE of 3.879%, showing superior performance and a more adequate adjustment to the seasonality and variability of the series.

The data processing carried out using R-Project ensured rigor in parameter estimation and in the comparison of results. The findings indicate that, although neural networks represent a modern and flexible alternative for time series forecasting, in this case the Box–Jenkins model offers greater precision and reliability. Thus, the results obtained constitute a relevant input for agricultural planning, resource management, and the formulation of policies aimed at strengthening the sugar industry in the country.

Keywords: Box Jenkins's methodology, multilayer perceptron, artificial neural network, SARIMA.

INDICE

DEDICATORIA	<i>i</i>
AGRADECIMIENTO	<i>ii</i>
PRESENTACIÓN	<i>iii</i>
RESUMEN	<i>iv</i>
ABSTRACT	<i>v</i>
ÍNDICE DE TABLAS	<i>vi</i>
ÍNDICE DE FIGURAS	<i>vii</i>
INTRODUCCIÓN	<i>1</i>
I PLANTEAMIENTO DEL PROBLEMA	<i>3</i>
1.1. Situación problemática	<i>3</i>
1.2. Formulación del problema	<i>4</i>
1.2.1. Problema general	<i>4</i>
1.2.2. Problemas específicos	<i>4</i>
1.3. Justificación de la investigación	<i>4</i>
1.4. Objetivos de la investigación	<i>5</i>
1.4.1. Objetivo general	<i>5</i>
1.4.2. Objetivos específicos	<i>5</i>
II MARCO TEÓRICO	<i>6</i>
2.1. Bases teóricas	<i>6</i>

2.1.1. Red neuronal	6
2.1.1.1. Características de una red neuronal.	9
2.1.2. Red Neuronal Artificial (RNA)	12
2.1.2.1. Estructura de una red neuronal artificial	14
2.1.2.2. Función de activación	15
2.1.3. Perceptrón	20
2.1.3.1. Perceptrón multicapa	21
2.1.3.2. Aprendizaje del perceptrón	23
2.1.4. Modelo de red neuronal artificial	24
2.1.5. Serie de tiempo	29
2.1.5.1. Variable aleatoria	30
2.1.5.2. Características de las series temporales	33
2.1.6. Metodología Box Jenkins	35
2.1.6.1. Estacionariedad	36
2.1.6.2. Transformación Box Cox	38
2.1.6.2.1. Transformación inversa	39
2.1.6.3. Autocorrelación	39
2.1.7. Modelo de serie de tiempo	41
2.1.7.1. Ruido blanco	41
2.1.7.2. Paseo aleatorio	42
2.1.8. Modelo ARIMA	42
2.1.9. Modelo SARIMA	54
2.1.9.1. Estimación de parámetros	56

2.1.9.2. Criterio de evaluación para la predicción	57
2.2. Marco conceptual	59
2.2.1. Caña de azúcar (<i>Saccharum officinarum</i> L.)	59
2.2.2. Aptitud del uso del suelo para el cultivo de caña de azúcar	61
2.2.3. Azúcar (Sacarosa)	62
2.3. Antecedentes empíricos	65
2.3.1. Antecedentes internacionales	65
2.3.2. Antecedentes nacionales	67
III HIPÓTESIS Y VARIABLES	69
3.1. Hipótesis	69
3.2. Hipótesis general	69
3.3. Hipótesis específicas	69
3.4. Identificación de variables e indicadores	69
3.6. Operacionalización de variables	70
IV METODOLOGÍA	71
4.1. Ámbito de estudio	71
4.2. Tipos y nivel de investigación	71
4.2.1. Tipo de investigación	71
4.2.2. Nivel de investigación	71
4.3. Población de estudio	71
4.4. Técnicas de recolección de datos	71

4.5. Técnicas de análisis e interpretación de la información _____	72
4.6. Técnicas para demostrar la verdad o falsedad de las hipótesis planteadas _____	72
<i>V RESULTADOS</i> _____	73
5.1. Análisis, interpretación y discusión de resultados _____	73
5.1.1. Metodología Box Jenkins _____	73
5.1.1.1. Estacionariedad _____	77
5.1.1.2. Características que presenta la serie _____	77
5.1.1.3. Estimación del modelo _____	83
5.1.1.4. Evaluación del modelo _____	84
5.1.1.5. Predicción del modelo _____	85
5.1.2. Método de red neuronal artificial _____	88
5.1.3. Comparación de método Box Jenkins y RNA _____	91
<i>DISCUSIÓN</i> _____	93
<i>CONCLUSIONES</i> _____	95
<i>RECOMENDACIONES</i> _____	96
<i>REFERENCIAS</i> _____	97
<i>ANEXOS</i> _____	103
a. Matriz de consistencia _____	103
b. Otros _____	107

ÍNDICE DE TABLAS

Tabla 1: Países con mayor producción de azúcar en el mundo _____	64
Tabla 2: Operacionalización de variables _____	70
Tabla 3: Cuadro resumen de estadísticos de la producción de azúcar en el Perú _____	73
Tabla 4: Test Dickey Fuller para probar la estacionariedad de la serie _____	77
Tabla 5: Test Mann Kendall para probar la tendencia de la serie _____	77
Tabla 6: Test Kruskal - Wallis para probar la estacionalidad de la serie _____	78
Tabla 7: Test Lagrange Multiplier para probar la variabilidad de la serie _____	79
Tabla 8: Comparación de modelos tentativos para la serie producción de azúcar _____	81
Tabla 9: Estimación de coeficientes para el modelo seleccionado _____	83
Tabla 10: Comparación del pronóstico de la producción de azúcar la data de prueba tomando en cuenta la transformación por Box Cox _____	85
Tabla 11: Comparación del pronóstico transformado con la data de prueba, producción de azúcar _____	87
Tabla 12: Comparación de modelos de redes neuronales artificiales _____	88
Tabla 13: _____	89
Tabla 14: Medidas de ajuste del método Box Jenkins y la red neuronal artificial _____	91

ÍNDICE DE FIGURAS

Figura 1: Partes del cerebro humano	7
Figura 2: La neurona y sus partes	8
Figura 3: Representación gráfica de la arquitectura del perceptrón multicapa	15
Figura 4: Gráfico de la función escalón	17
Figura 5: Gráfico de la función lineal y mixta	18
Figura 6: Gráfico de la función sigmoideal	19
Figura 7: Red de dos capas	27
Figura 8: Gráfico de autocorrelación del modelo AR (2)	46
Figura 9: Gráfico de autocorrelación del modelo MA (2)	48
Figura 10: Gráficos de autocorrelación, ARMA (1,1)	60
Figura 11: Países productores de caña de azúcar	58
Figura 12: Comportamiento de la producción de azúcar en el tiempo (1994 –2023)	74
Figura 13: Comportamiento de los datos por mes en un diagrama de cajas	74
Figura 14: Comportamiento de los datos por años en un diagrama de cajas	74
Figura 15: Representación gráfica de la autocorrelación simple	79
Figura 16: Representación gráfica de la autocorrelación parcial	79
Figura 17: Pronóstico y valor ajustado por el modelo SARIMA(3,0,1)(0,1,1) ₁₂ para la serie de producción de azúcar	85
Figura 18: Pronóstico y data de prueba de la serie producción de azúcar	86
Figura 19: Predicción y valor ajustado por la RNA para la producción de azúcar	89
Figura 20: Comparación de la predicción con la data de prueba, producción de azúcar	89
Figura 21 Gráfico de comparación SARIMA, RNA y los datos de prueba del año 2023	90

INTRODUCCIÓN

El azúcar es considerado una fuente muy importante de energía para el cuerpo humano que permite recuperar las fuerzas de manera rápida, además de ser considerado como un ingrediente fundamental en la alimentación, esto por aportar sabor dulce, es un edulcorante (endulza alimentos), del cual se puede disfrutar, (Comité Nacional para el Desarrollo Sustentable de la caña de Azúcar, 2016). El azúcar es elaborado a partir de caña de azúcar, que es un producto cultivado, por esto las empresas azucareras generan más de 100 000 empleos, de manera directa e indirecta, conllevando así a más de 200 000 personas dependientes de este producto, tanto para subsistir como para consumir (PERUCAÑA, 2023); a pesar de los beneficios que aporta es importante consumir con moderación porque en exceso puede generar enfermedades.

El sector del azúcar en el Perú ha mostrado un progreso sostenido en las últimas décadas, incorporando innovaciones tecnológicas que mejoran la productividad y la competitividad. De igual manera, la investigación académica ha evolucionado en el uso de modelos de predicción, pasando de enfoques tradicionales como Box Jenkins a herramientas modernas como las redes neuronales artificiales. Este avance refleja la importancia de aplicar metodologías más robustas para comprender y anticipar el comportamiento de la producción de azúcar, lo que fortalece la toma de decisiones en un sector estratégico para la economía nacional.

Por esto, la importancia de tener herramientas adecuadas de estadística, como la metodología de Box Jenkins y las redes neuronales artificiales (RNA), para realizar predicciones sobre la producción de azúcar, contribuiría a planificar y optimizar los recursos humanos y financieros para prever el comportamiento futuro de la producción de azúcar.

La estructura del proyecto de investigación se divide en cinco capítulos, y están detallados de la siguiente manera:

El primer capítulo contiene el planteamiento del problema, la formulación del problema, la justificación del problema y los objetivos de la investigación.

El segundo capítulo abarca el marco teórico, en el cual se exponen las bases teóricas, el marco conceptual y los antecedentes.

En el tercer capítulo se menciona las hipótesis, identificación y operacionalización de las variables e indicadores.

El cuarto capítulo es sobre la metodología, el ámbito de estudio, tipo y nivel de investigación, la población de estudio, técnicas de recolección y análisis de datos y las técnicas para demostrar las hipótesis.

En el capítulo cinco se muestran los resultados, mediante gráficos y tablas son las interpretaciones correspondientes.

Por último, se presenta la discusión, conclusiones y las recomendaciones, seguida de las referencias y anexos.

I PLANTEAMIENTO DEL PROBLEMA

1.1. Situación problemática

La producción de azúcar en el Perú constituye una de las actividades agrícolas e industriales más relevantes, pues no solo abastece la demanda interna de consumo, sino que también sostiene una cadena productiva con alto impacto social y económico.

La producción de azúcar en las diferentes empresas azucareras genera 25 000 empleos directos y 125 000 empleos indirectos, además más de 100 000 personas adicionales aproximadamente mantienen un vínculo con productos de la agroindustria del azúcar, del cual se puede indicar que más de 200 000 personas dependen de este sub sector económico (Asociación Peruana de Productores de Azúcar, 2002). Mas de 20 000 productores de caña de azúcar y más de 100 000 empleos se mantiene a la actualidad con la producción de azúcar, menciona (PERUCAÑA, 2023) en su informe anual del 2023, recalcando cinco departamentos productores (Lima, Áncash, La Libertad, Lambayeque y Piura) en el norte del Perú.

Históricamente, la producción de azúcar en el Perú ha enfrentado altibajos como la modernización desigual de los ingenios, los cambios en los precios internacionales, la variabilidad climática y la competencia con otros productos agroindustriales.

La desaceleración económica mundial y el esfuerzo para contrarrestar la inflación generada por la pandemia COVID – 19 se vio marcado en el año 2023, y el Perú no fue la excepción generando una contracción del 0.6% del PBI, de esta manera afectando a la producción de azúcar, además de la inestabilidad política, la inseguridad y los factores climáticos contribuyendo en la desestabilización económica y productiva. (PERUCAÑA, 2023)

En este contexto, comprender y anticipar el comportamiento de la producción de azúcar resulta esencial para una adecuada planificación y optimización de los recursos humanos y financieros, y diseños de estrategias de competitividad. Por ello, se propone analizar modelos de Box – Jenkins y redes neuronales artificiales con el fin de identificar cuál ofrece mayor precisión para predecir la producción de azúcar y así contribuir al fortalecimiento de este sector estratégico.

1.2. Formulación del problema

1.2.1. Problema general

¿Cuál de los modelos según la metodología Box Jenkins o las redes neuronales artificiales permite una mejor predicción de la producción de azúcar?

1.2.2. Problemas específicos

¿Cuál es la predicción de la producción de azúcar mediante modelos de series de tiempo utilizando la metodología de Box Jenkins?

¿Cuál es el pronóstico para la producción de azúcar utilizando modelos de redes neuronales artificiales?

¿Cuál de los modelos realiza una mejor predicción para la producción de azúcar utilizando redes neuronales artificiales o la metodología Box Jenkins?

1.3. Justificación de la investigación

La presente investigación muestra el uso de modelos de la metodología Box – Jenkins y redes neuronales artificiales aplicadas a la producción de azúcar para mostrar el pronóstico de la producción futura de azúcar, en base a los datos pasados de producción, y de esta forma probar la utilidad que aportan como modelos de predicción.

Factores ambientales, así como la pandemia COVID – 19, afectaron la producción de azúcar estos últimos años, esto se confirma con la producción menor que fue en el año 2023 comparado con el año 2022, que de acuerdo a la empresa productora de azúcar Sol de Laredo, la producción de azúcar nacional en el año 2023 fue de 1 062 623 toneladas, 4.4% menos comparado al año 2022, bajo estos datos podemos suponer que, si esto continua de esta forma, podría afectar la economía del país, lo que conlleva que busquemos prever la producción futura y buscar soluciones a posibles problemas con anterioridad.

El azúcar destaca un rol importante en el cuerpo humano, su función es producir energía para el funcionamiento de diferentes órganos como el cerebro, que consume el 20% de energía del cuerpo y la glucosa es su único sustrato, los tejidos también lo necesitan, con un consumo medible y no en exceso. (Ministerio de Agricultura, Pesca y Alimentación, s.f.)

1.4. Objetivos de la investigación

1.4.1. Objetivo general

Analizar los modelos de la metodología Box Jenkins o la red neuronal artificial para encontrar la predicción adecuada de la producción de azúcar en el Perú.

1.4.2. Objetivos específicos

Determinar la predicción de la producción de azúcar utilizando modelos de series de tiempo mediante la metodología de Box Jenkins.

Determinar el pronóstico de la producción de azúcar utilizando modelos de redes neuronales artificiales.

Comparar y evaluar la mejor predicción de los modelos de redes neuronales artificiales o la metodología Box Jenkins.

II MARCO TEÓRICO

2.1. Bases teóricas

2.1.1. *Red neuronal*

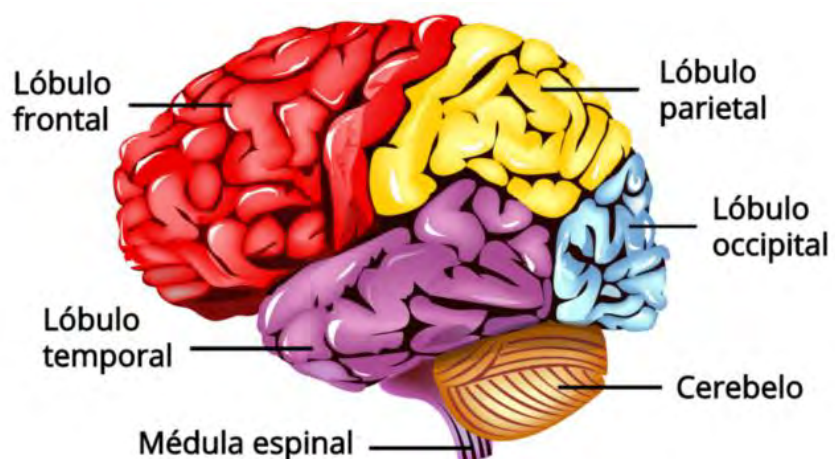
Según Hilera Gonzales y Martínez Hernando (1995), una red neuronal o computación neuronal es una inspiración de redes neuronales biológicas, células neuronales conocidas como “neuronas”, los cuales, mediante interconexiones a través de las ramas de salida, que son los axones, producen conexiones con las demás neuronas (sinapsis), en el cual se puede encontrar la complejidad por la abundancia de lazos que presentan.

El arduo trabajo de los investigadores a lo largo de los años, donde realizaron experimentos con el fin de poder comprender la inteligencia humana que abarca una complejidad única, dio sus frutos conllevando a una investigación ahora útil, usando las redes neuronales biológicas para crear las redes neuronales artificiales.

Russell (2018), menciona que las investigaciones se enfocaron en enfermedades y heridas para comprender como el cerebro controla dichos movimientos y otras funciones primordiales, a pesar de la información limitada, el hallazgo al entender las partes del cerebro relacionados a ciertas lesiones específicas y las funciones que cada lado del cerebro cumplía, como se muestra en la Figura 1.

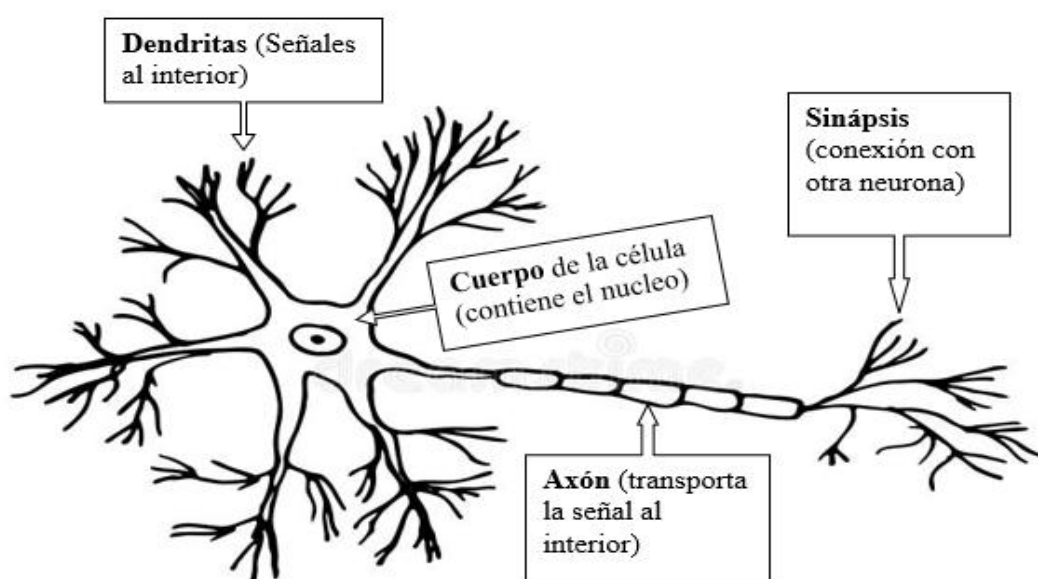
Figura 1

Partes del cerebro humano



Fuente: Enciclopedia humanidades (2023).

En cierta forma, sin darnos cuenta realizamos una construcción de una red neuronal, al tratar de comprender lo que nos rodea, tratar de arreglar algo que está roto o comprender el funcionamiento de algunos aparatos abriendo y observando lo que tienen adentro por simple y espontanea curiosidad, en pocas palabras podemos decir que una red neuronal es un sistema que hace cálculos.

Figura 2*La neurona y sus partes*

Fuente: Hilera Gonzales y Martínez Hernando (1995).

La Figura 2 muestra la representación gráfica de una neurona biológica, con sus partes que en una red neuronal se trataron de replicar, por lo menos en el funcionamiento. La característica que diferencia de las demás células, es que las neuronas tienen la capacidad de comunicarse, las **dendritas** reciben las señales de entrada, se combina e integra en el cuerpo de la célula y emite una señal de salida que es transportada mediante el **axón** hacia otra neurona, distribuyéndose mediante la **sinápsis**; cabe referir que se estima un aproximado de 10^{15} conexiones donde una neurona recibe y envía miles de señales con información. (Hilera Gonzales & Martínez Hernando, 1995)

La función de una neurona tanto en animales y personas es, mediante la conexión de los órganos de los sentidos con los músculos y glándulas, recoger, transmitir y almacenar la información para luego elaborar una nueva información y enviarla hacia otras neuronas, en su mayoría estas informaciones enviadas lo hacen en forma de redes, donde las receptoras

realizan el mismo procedimiento, claro, tomando en cuenta su estado de activación que le permite realizar dichas funciones. Debido a sus millones de conexiones la dificultad de crear un dispositivo con semejantes características es sumamente complicado o imposible que marca una limitación en el sistema artificial. (Isasi Viñuela & Galvan León, 2004)

Las redes neuronales presentan muchas ventajas, desde su capacidad de aprender a partir de experiencias, organizar la información, tolerancia a fallos reteniendo información a pesar de sufrir grandes daños, su operación es en tiempo real además de la facilidad de inserción en la tecnología actual para mejorar la capacidad y realizar tareas. (Hilera Gonzales & Martínez Hernando, 1995)

2.1.1.1. Características de una red neuronal.

a. Por su topología

Esta característica se basa tomando en cuenta su organización y en cómo se disponen las neuronas en las capas.

- **Red monocapa:** Las conexiones entre neuronas se hacen en una sola capa. Por ejemplo, la red de Hopfield, red de Brain-State-in-a-Box donde hay conexión autorrecurrentes, etc.
- **Redes multicapa:** Las conexiones que se da entre las neuronas son en varias capas o niveles, en el cual se encuentran conexiones hacia adelante (feedforward) o conexiones hacia atrás (feedback), por ejemplo, las redes feedforward como Perceptrón, Adaline, Madaline, etc. y redes feedback como ART (Adaptive Resonance Theory) y BAM (Bidirectional Associative Memory) que son las más conocidas. (Hilera Gonzales & Martínez Hernando, 1995)

b. Por su aprendizaje

Es el proceso en el cual se modifica sus pesos de acuerdo a la información de entrada en la conexión de neuronas. Si los pesos son distintos de cero entonces implica la creación de una nueva conexión mientras que si el peso o pesos son iguales a cero es cuando se destruyen las conexiones, además de concluir que la red a aprendido y los pesos son estables.

Ahora el detalle es cómo modificar estos pesos para que el modelo aprenda, por esto se ha determinado una regla de aprendizaje, que consta en tomar en cuenta la existencia o no de un supervisor y de un aprendizaje “*on line*” (la red aprende mientras se encuentra en funcionamiento) u “*off line*” (la red se desconecta hasta que termine su procesamiento) en esta última fase se puede distinguir dos conjuntos de datos como son de entrenamiento y prueba, a ser utilizados para su análisis correspondiente, además de que sus pesos permanecen fijos y de esa manera evitando problemas de estabilidad en su funcionamiento, lo cual no se puede apreciar en la primera fase. (Hilera Gonzales & Martínez Hernando, 1995)

- **Redes neuronales con aprendizaje supervisado:** El proceso donde aprende se hace a través de un entrenamiento controlado por un agente externo, de tal forma se comprueba la salida y de no coincidir se modifica los pesos hasta que el resultado sea lo adecuado o se aproxime a lo deseado.

Hilera Gonzales y Martínez Hernando (1995), clasifican en tres tipos de aprendizajes supervisados.

Aprendizaje por corrección de error, consiste en ajustar los pesos en función a la diferencia de los valores que se tiene como prueba y los resultados obtenidos.

Aprendizaje por refuerzo, se considera el más lento que el anterior debido a que no desea que se le dé la información deseada completa para que aprenda, donde la función del supervisor es indicar mediante señal si la salida de resultados es lo que se quiere o no (éxito = +1, fracaso = -1) que mediante probabilidades, tomando en cuenta la información obtenida, ajusta los pesos, los autores comparan este tipo de aprendizaje con un crítico debido a que el supervisor hace la función de opinar sobre los resultados y no indica lo que debe salir.

Aprendizaje estocástico, realiza cambios aleatorios en cuanto a los valores de los pesos evaluando su efecto con respecto a lo deseado, si el comportamiento se acerca al objetivo deseado se acepta dichos cambios, en caso contrario se toma en cuenta una determinada distribución de probabilidades para ser aceptada.

- **Redes neuronales con aprendizaje no supervisado:** Conocido como auto supervisado debido a que no necesita de un agente externo para ajustar sus pesos, tampoco recibe información sobre la salida deseada, y esto conlleva a su capacidad de organizarse por su cuenta.

Hilera Gonzales y Martínez Hernando (1995), mencionan dos tipos de aprendizaje no supervisado.

Aprendizaje Hebbiano, consiste en ajustar los pesos con la correlación de los valores de activación o salida de las neuronas conectadas, es decir que extrae las características de los datos de entrada.

Aprendizaje competitivo y cooperativo, donde las neuronas compiten y cooperan entre todas para completar una tarea asignada, en el proceso compiten para activarse y obtener una respuesta óptima, quedando así una neurona de salida mientras que las demás quedan anuladas.

c. Por la asociación de información entre la entrada y salida

Las redes neuronales tienen la capacidad de almacenar la información que aprenden, la cual se distribuye mediante los pesos, que de acuerdo a ello muestran una respuesta que viene a ser la salida.

Hilera Gonzales y Martínez Hernando (1995), clasifican en dos formas o tipos según el modo que usan para aprender.

- **Red heteroasociativa**, quien requiere de al menos dos capas para captar y retener la información que entra y la otra para la salida quien debe haber relación con la información de entrada, su modo de aprendizaje es en pareja, es decir si se introduce una información A_i la salida será una información B_j , esto porque al aprender lo hace de la forma $(A_1, B_1), \dots, (A_n, B_n)$.

- **Red autoasociativa**, a diferencia de lo anterior, esta red puede o funciona con una sola capa de neuronas, quien retiene la información de entrada y muestra una salida autoasociada, es decir que al ser ingresado una serie de datos A_i realizará una autocorrelación y comparando con un dato almacenado que sea lo más parecido a la información de entrada. Este tipo de red se utiliza para filtrar información o una reconstrucción de datos sin distorsión o ruido, y aprovechar su autocorrelación para buscar información o contenido relacionado en una base de datos. (Hilera Gonzales & Martínez Hernando, 1995)

2.1.2. Red Neuronal Artificial (RNA)

Isasi Viñuela y Galvan León (2004), afirman que sus inicios datan a principios del siglo XIX, empezado por Freud que en los años 40 en adelante cogió relevancia, en dicha década Warren McCulloch y Walter Pitts encontraron el primer modelo matemático, las neuronas operan a través de impulsos binarios como también hay un umbral; Donald Hebb

desarrolló el procedimiento del aprendizaje matemáticamente cuyo nombre lleva aprendizaje Hebbiano, ya en el año 1951 se obtuvo los resultados iniciales en la práctica por Marvin Minsky y de esta manera nació las redes neuronales artificiales.

Hilera Gonzales y Martínez Hernando (1995), mencionan que las redes neuronales artificiales muestran características semejantes al cerebro humano, con ello refieren que es capaz de aprender de la experiencia, abstraer características primordiales a partir de entradas que muestran información irrelevante, etc. De esta forma se puede observar las numerosas ventajas que ofrece, se adapta de acuerdo a su aprendizaje, puede organizar la información que recibe, a pesar de sufrir una destrucción parcial de la red que conduce a su degradación de su estructura se puede retener algunas capacidades de la red, operación en tiempo real y una fácil inserción en la tecnología actual.

De acuerdo a Russell (2018), el modelar un RNA simple con un sistema electrónico simple es fiable además de ser bastante fácil, por ejemplo, el algoritmo usado en las computadoras que hacen que nos muestren sus actividades. El descubrimiento fundamental en la investigación realizada por los científicos, fue sobre el proceso por la cual una neurona pasa una señal a otra neurona, notaron que, al momento de procesar información, las grandes y complicadas células son las más importantes debido a que se encargan de la comunicación entre las neuronas.

Lo que se puede recopilar de esto como una ventaja es que las RNAs tienen estructuras fáciles de trazar para que puedan ser utilizadas además de que su producción es económica, el objetivo que poseen es bien definido y están enlazados con una estructura simple, pero hay factores a una red neuronal artificial:

1. Elementos usados para la construcción de la red (como se ven y funcionan las neuronas).
2. El cómo de conectar las neuronas.
3. Establecimiento de los argumentos y parámetros de la red mediante el proceso de aprendizaje.

Es importante conocer que en una RNA cada neurona cuenta con una memoria interna específica, que se puede representar como pesos y sesgos, además de contar ciertas habilidades para convertir señales de entrada en señales de salida. (Russell, 2018)

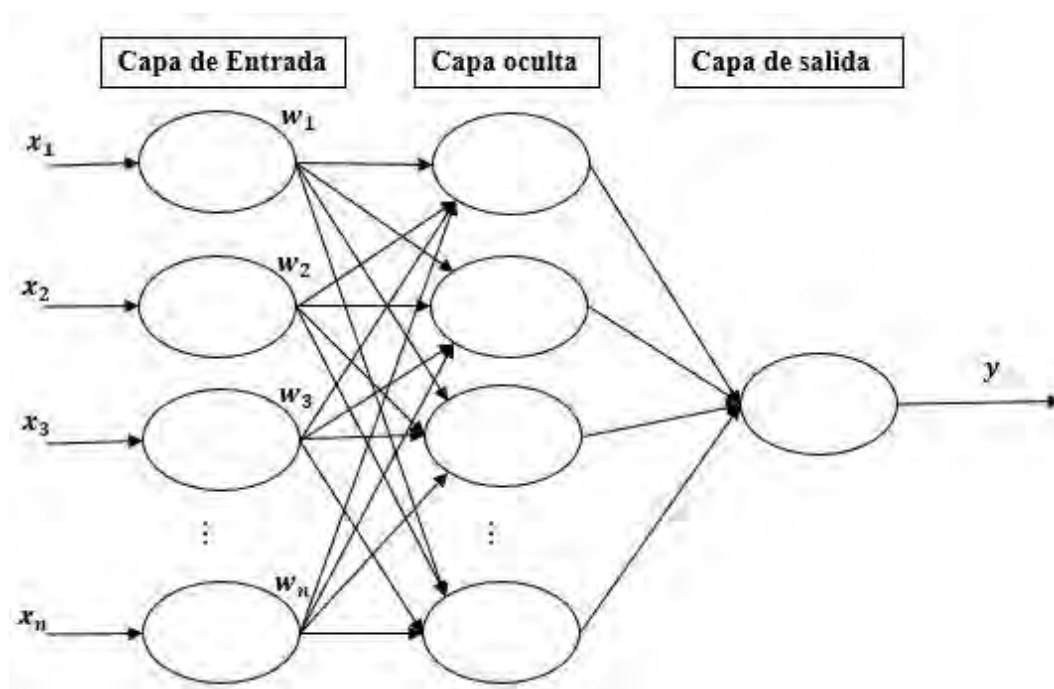
2.1.2.1. Estructura de una red neuronal artificial

Hilera Gonzales y Martínez Hernando (1995), indican que para la distribución de los nodos o neuronas dentro de la red se clasificaron en nivel o capas, los cuales cuentan con un número determinado de neuronas, esta parte por una capa de entrada, nivel por el cual se introduce o recibe la información proporcionada desde el exterior, esta pasa a una capa oculta, donde se realiza el procesamiento de los datos, que va desde su aprendizaje hasta su almacenamiento, dentro de este nivel podría haber cero o más subniveles con grupos de neuronas interconectadas de diferentes formas, y por último se halla la presencia de una capa de salida por el cual se transfiere el resultado hacia el exterior.

En su investigación, Isasi Viñuela y Galvan León (2004), afirman sobre la estructura de la red, agregando que la función de la capa de entrada cumple con el rol de propagar la información que recibe, en forma de señal, a todas las neuronas ubicadas en la capa oculta quien se encarga del procesamiento terminando en un resultado que sale por la última capa, esto se puede apreciar en la Figura 3.

Figura 3

Representación gráfica de la arquitectura del perceptrón multicapa



Redes feedforward o redes alimentadas hacia adelante con conexiones de pesos entre las neuronas son características mostrada por un perceptrón multicapa, además de contar con un umbral que es como una conexión más a la neurona, una entrada constante e igual a 1. (Isasi Viñuela & Galvan León, 2004)

2.1.2.2. Función de activación

Hilera Gonzales y Martínez Hernando (1995), explican que en cada unidad i -ésima, cada neurona en cualquier instante se encuentran caracterizadas por un estado de activación ($a_i(t)$), y asociado a cada unidad se encuentra una función de salida (f_i) quien se encarga de transformar el estado de activación actual en una señal de salida (y_i), los cuales son enviadas hacia otras neuronas de la red, donde de acuerdo con la sinapsis cada señal es modificada (peso, w_{ij}), esto siguiendo una determinada regla, de esta manera las señales llegan a una unidad j -ésima generando una combinación denominada la entrada total (Net_j), aquí una

nueva función de activación (F), determinará un nuevo estado de activación ($a_j(t+1)$), siempre tomando en cuenta la entrada total y el estado de activación anterior $a_j(t)$.

Isasi Viñuela y Galvan León (2004), postulan en su investigación que una neurona artificial tiene un estado interno conocido nivel de activación, lo que le permite cambiar de estado, es decir que puede estar en un estado activo o inactivo.

La información llega de manera continua a cada neurona, ellas se encargan de evaluar sus estados continuamente, de manera independiente, de acuerdo a la llegada de la información (modo asíncrono), o de manera simultánea como si hubiera un reloj decidiendo cuando cambiar su estado. Toda neurona de una red se encuentra en un estado (reposo o excitado), los valores de activación pueden ser discretos (conjunto pequeño de valores o valores binarios como 0 y 1) o continuos (se le asigna valores $[0, 1]$ o $[-1, 1]$), limitados o ilimitados, generalmente siguen una función sigmoideal. Es necesario conocer las reglas a seguir para que una neurona logre el estado de activación, una primera forma es en conjunto, es decir que, dependiendo a la fuerza, peso o magnitud que produce una neurona sobre otra se dará su estado de activación con la iteración entre las neuronas, y una segunda forma es cuando la señal enviada a cada neurona y estas a sus vecinas dependerá de cada estado de activación que poseen, es decir de manera individual. (Hilera Gonzales & Martínez Hernando, 1995)

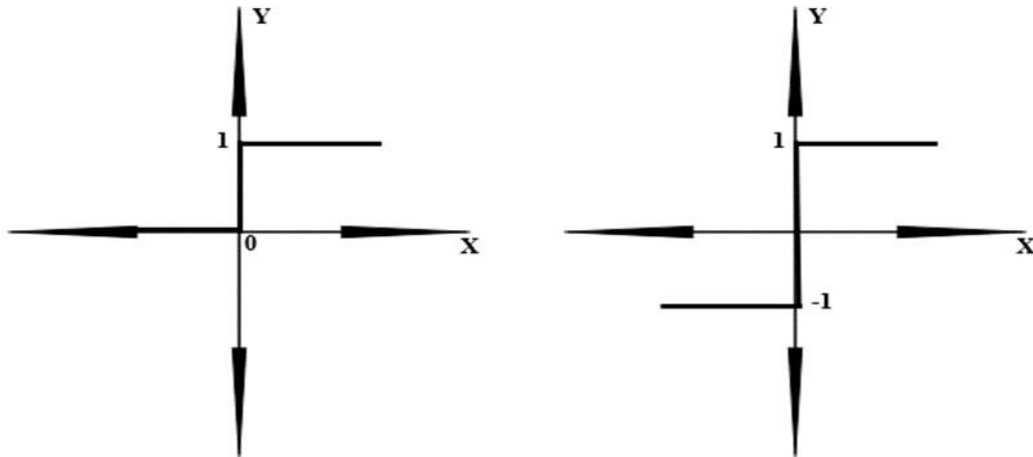
Función de activación escalón

Esta función de activación es considerada la más fácil porque se asocia a neuronas binarias, si la suma de las entradas es mayor o igual al umbral de la neurona entonces la activación es 1, caso contrario es 0 o podría ser -1, a pesar de la facilidad de uso y lo fácil

que es al implementar en hardware, se ve limitado en cuanto a sus capacidades, como se observa en la Figura 4. (Hilera Gonzales & Martínez Hernando, 1995)

Figura 4

Gráfico de la función escalón



Nota. El umbral se toma 0, caso contrario el escalón quedaría desplazado. Fuente: Hilera Gonzales y Martínez Hernando (1995).

Cuya ecuación para la activación en una red neuronal es,

- Cuando el estado de activación $E = \{0,1\}$

$$y_i(t+1) = \begin{cases} 1 & \text{si } Net_i > \theta_i \\ y(t) & \text{si } Net_i = \theta_i \\ 0 & \text{si } Net_i < \theta_i \end{cases}$$

Con " Net_i ", es la suma total del producto de cada señal por el peso y " θ_i " representa al umbral.

- Cuando el estado de activación $E = \{-1,1\}$

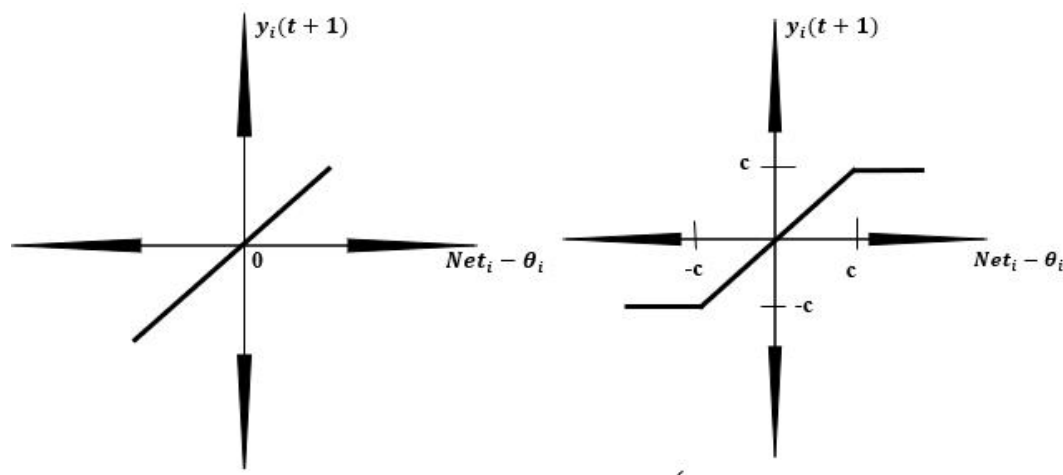
$$y_i(t+1) = \begin{cases} 1 & \text{si } Net_i > \theta_i \\ y(t) & \text{si } Net_i = \theta_i \\ -1 & \text{si } Net_i < \theta_i \end{cases}$$

Función de activación lineal y mixta

Conocida también como función identidad cuya expresión es $f(x) = x$, si la suma de señales de entrada es menor que un límite inferior la activación es 0 o -1, si es mayor o igual que el límite superior entonces la activación es 1, y si la suma se encuentra entre el límite superior e inferior entonces la activación es una función lineal. La representación gráfica se observa en la Figura 5. (Hilera Gonzales & Martínez Hernando, 1995)

Figura 5

Gráfico de la función lineal y mixta



Nota. Límite superior (c), límite inferior (-c) con ciclo de operación (x). Fuente: Hilera Gonzales y Martínez Hernando (1995).

Para la función lineal o identidad, el estado de activación puede tomar cualquier número de la recta real, cuya ecuación es:

$$y_i(t+1) = Net_i - \theta_i$$

Para la función lineal-mixta, su ecuación,

$$y_i(t+1) = \begin{cases} c & Net_i \geq c \\ Net_i - \theta_i & (-c) + \theta_i < Net_i < c + \theta_i \\ -c & Net_i \leq (-c) + \theta_i \end{cases}$$

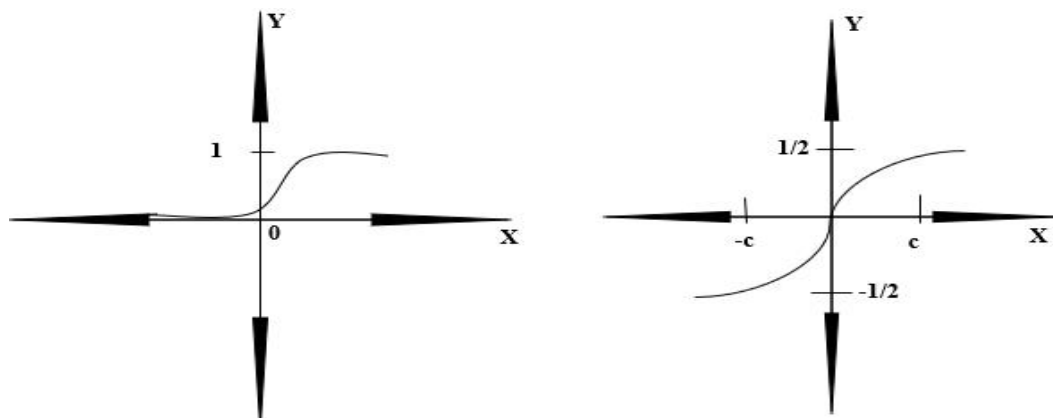
Esta función hace que la función de activación se encuentre obligado a estar dentro del intervalo de valores reales prefijados.

Función de activación sigmoideal

La importancia de la función sigmoideal radica en que su derivada siempre es positiva y cercana a 0, ya sean valores positivos o negativos incluso si son grandes, si $x = 0$ toma su valor máximo, esto ayuda que se puedan utilizar reglas de aprendizaje, basta con que tenga límites superiores e inferiores podrá realizar una activación de manera satisfactoria, donde el valor de salida se encuentra comprendido en la zona alta o baja del sigmoide, la representación gráfica de la función sigmoideal se aprecia en la Figura 6. (Hilera Gonzales & Martínez Hernando, 1995)

Figura 6

Gráfico de la función sigmoideal



Nota. Ciclo de operación (x) y suma ponderada de las entradas más el sesgo (a).

Fuente: Hilera Gonzales y Martínez Hernando (1995).

$$f(x) = \frac{1}{1 + e^{-ax}}$$

Para asegurar la salida de la información de una neurona determinada, se considera que hay una neurona ficticia, cuyo valor de salida es la unidad, asociado a la neurona en la entrada mediante el peso.

La conexión mediante la sinapsis que se da de una neurona a otra, digamos que de una neurona i a una neurona j para diferenciarlos, esta ponderada mediante un peso w_{ij} , el cual mide el impacto que la neurona i tiene sobre la neurona j , además pueden ser negativos, positivos o nulos, y al conjunto o matriz de los pesos se denomina \mathbf{W} , en el cual afectará en la interacción entre ambas neuronas, si es positivo la interacción es excitadora, esto indica que siempre que la neurona i se encuentre activada y la información recibida por la neurona j tenderá a activarla a esta última, si el peso es negativo entonces nos referimos a una interacción inhibidora, es decir que la información de la neurona i (activada) tenderá a desactivar a la neurona j , y finalmente cuando el peso es nulo es porque no hay interacción o conexión entre las neuronas. (Hilera Gonzales & Martínez Hernando, 1995)

Es importante tener en cuenta que biológicamente la información o el conocimiento se encuentra en la conexión entre las neuronas (sinapsis), pero en el caso de una red neuronal artificial el conocimiento se encuentra en los pesos, por ello, ya que el aprendizaje se basa en realizar cambios en las conexiones, es lógico referir que las redes neuronales van aprender si se modifica los valores de sus pesos, claro que también dependerá del número de neuronas que hay y como se conectan estas.

2.1.3. Perceptrón

Dentro de las características se mencionó el perceptrón, utiliza aprendizaje supervisado con conexión feedforward (hacia adelante) y trabaja con capas múltiples, y este tipo o modelo de red fue el primero de una red neuronal artificial. Rosenblat (1958, como se

citó en Hilera Gonzales y Martínez Hernando, 1995) fue quien desarrollo este modelo, lo cual en los años 60 despertó un gran interés en la comunidad por su capacidad de aprender a base de patrones sencillos como también reconociéndolos, claro que esos años se empezó con una función de activación lineal y escalón, y su capacidad de diferenciar la clase de entrada que se le introducía fue la más sorprendente.

El perceptrón se utiliza para clasificar y diferenciar entradas mediante una línea recta, lo que lo hace fácil de usar en problemas simples, de esta forma generando una limitación para problemas más complejos, además de que utiliza funciones de activación básicas y en la práctica los datos no se comportan linealmente por lo que dicha función de activación ya no es lo apropiado, por estas limitaciones es que nace una extensión denominada perceptrón multicapa resolviendo ciertas limitaciones.

2.1.3.1. Perceptrón multicapa

Isasi Viñuela y Galvan León (2004), mencionan que la agrupación de neuronas en diferentes capas que tiene es lo que caracteriza al perceptrón multicapa, en esta arquitectura se encuentra neuronas que se encargan solo de recibir señales del exterior y están en la capa de entrada, su función es propagar las señales a toda neurona que se encuentra en la capa oculta donde se realiza el procesamiento no lineal, y por último pasa por la capa de salida como respuesta para cada señal de entrada que hubo.

Russell (2018) menciona que la suma de las señales de entrada es multiplicada por los pesos, en el cual usa una función de transferencia no lineal con una forma sigmoide es referido a neuronas de tipo perceptrón.

El perceptrón multicapa es una generalización del perceptrón simple, donde Minsky y Papert en 1969, a partir de una combinación de perceptrones simples en las neuronas

ocultas lograron generar un perceptrón multicapa que resulto útil para abordar soluciones a ciertos problemas no lineales, pero faltó adaptar los pesos en la capa de entrada a la capa oculta, lo que inspiró a investigaciones futuras, en los cuales Rumelhart, Hinton y Williams en 1986 lograron presentar una forma de retro propagación de los errores medidos en la salida hacia las neuronas ocultas, conocido como la regla delta generalizada, y de esto, distintos autores demostraron que el perceptrón multicapa es un aproximador universal, esto porque cualquier función continua sobre un compacto de \mathbf{R}^n puede aproximarse a un perceptrón multicapa, por ello y por la facilidad de uso además de su aplicabilidad, dicha arquitectura es muy usada para resolver problemas en diferentes áreas como son reconocimientos de caracteres ópticos y escritos, predicción de series de tiempo, etc.; a pesar de todo cuenta con limitaciones, el largo proceso de aprendizaje que proviene de un buen número de variables, análisis con la presencia de componentes no lineales, esto podría reducir su precisión en algunas aplicaciones. (Isasi Viñuela & Galvan León, 2004)

En conclusión, el perceptrón cuenta con una sola capa de neuronas y su salida es prácticamente la suma ponderada más el sesgo con una función de activación del tipo escalón que cuenta con limitaciones como la clasificación de datos linealmente separables y por ello no es apropiado para problemas complejos, y debido a estas limitaciones se encuentra el perceptrón multicapa quien maneja capas múltiples y funciones de activación no lineales como el sigmoideal y relu (si el valor de entrada es menor a cero entonces la función de activación es cero caso contrario es el mismo valor), de esta forma resolviendo sus limitantes del perceptrón simple, en ambos casos su aprendizaje es supervisado, lo que indica que los resultados son evaluados y modificados si se encuentra necesario.

2.1.3.2. Aprendizaje del perceptrón

Esta es la parte más interesante e importante de una red neuronal, ya que depende de esto el tipo de problemas que puede resolver o es capaz de hacerlo, y en las redes neurales artificiales el aprendizaje se da a partir de ejemplos entonces para que este sea óptimo se debe dar un número suficiente de ejemplos, caso contrario la red no aprenderá lo que conlleva a un ajuste ineficaz de pesos, a esto Isasi Viñuela y Galvan León (2004) refieren como una característica que debe poseer la red para su aprendizaje y lo denominan que debe ser significativo, así como representativo, esto indica que al introducir ejemplos para su aprendizaje, se debe aplicar a todas las regiones, de lo contrario, si se encuentra un conjunto con más ejemplos de aprendizaje, la red tomará en cuenta más esa área y no al resto, conllevando a un aprendizaje incompleto.

Russel (2018) menciona algo a tener en cuenta, y es, que al añadir peso a una estructura de neurona hace que esta obtenga la capacidad de aprender, además de que se puede agregar señales de entrada ya después de introducir señales anteriores, esto se conoce como estimulación acumulativa o postsináptica porque se añade un componente extra, a esto se le denomina como un sesgo que también pasa por los métodos de aprendizaje, que a su vez puede considerarse como un peso extra asociado a las entradas que brinda señal interna con un valor 1 que es constante. Al ser introducida la señal, hay veces en las que es enviada directamente al axón y considerada como una señal de salida lo cual funciona en sistemas lineales.

Para que la red neuronal aprenda son importantes los pesos, de esto dependerá su aprendizaje eficiente, el proceso viene de introducir los ejemplos o datos y de acuerdo a un esquema de aprendizaje ir modificando los pesos, esto se puede hacer cada vez que se

introduce o al termino de los ejemplos introducidos, hasta que cumpla cierto esquema de convergencia.

Los criterios de convergencia para que se detenga el proceso de aprendizaje van desde determinar un número fijo de ciclos con anterioridad, establecer un error ya sea a nivel individual o general que si la red produce un valor menor a éste parará su proceso, pero esto ocurre con menor frecuencia en la práctica lo que conlleva a depender de un ciclo prefijado, y por último definir un esquema de aprendizaje que haga a las conexiones modificarse con menor intensidad hasta alcanzar un estado en el que no converja y se detenga, pero esto conlleva tres tipos, el aprendizaje supervisado en el que se hará una comparación de la salida con los valores deseados, si la diferencia es mucho se puede modificar los pesos para acercar dicha diferencia, luego viene el aprendizaje no supervisado quien no necesita de un agente externo y realizando todo ajuste internamente dependiendo de los valores de entrada denominándose por esto sistema autoorganizada, y por último se tiene el aprendizaje por refuerzo que es una variante del aprendizaje supervisado con la diferencia de que solo se verifica la salida, sin tomar en cuenta el error que cometió la red para cada ejemplo pero si modificando los pesos en busca de salidas adecuadas (Isasi Viñuela & Galvan León, 2004), para ello es necesario dividir los datos en dos grupos, entrenamiento y prueba.

2.1.4. Modelo de red neuronal artificial

La red neuronal artificial se caracteriza por tener una o más entradas, pero solo por una salida (capa de entrada, capa oculta y capa de salida), esto mediante información que acabará en un resultado específico.

Las neuronas reciben entradas de señales, X_i , en base a ello determina una respuesta con una sola señal de salida, W_i es un parámetro peso cuya función es separar las entradas de

la neurona determinando el grado de importancia en la información que llega a una neurona mediante una entrada específica, x_i , si en caso llega una señal en particular por una entrada, esta es modificada por el uso del peso de la entrada, esta modificación a menudo está basada en la multiplicación por el peso que posteriormente la señal participa de forma distinta, otro dato a tener en cuenta es si el peso es mayor o menor que 1 y de acuerdo a ello la señal es bastante fuerte (si valor peso mayor que 1), además de ser conocidas como entradas inhibitorias si los pesos son negativos, y aquellos con pesos positivos tienden a ser llamados entradas excitantes. (Russell, 2018)

La salida producida de una neurona, denominada N , se puede definir como:

$$N = x_1w_1 + x_2w_2 + \dots + x_nw_n$$

O en su forma vectorial,

$$N = \mathbf{X}^T \mathbf{W}$$

Estas señales de salida son procesadas por “F”, una función de activación o salida, cuyos valores pasan por una evaluación de los pesos que se ajustan durante el aprendizaje. (Isasi Viñuela & Galvan León, 2004)

Las neuronas artificiales introducen y producen señales, tanto de entrada y salida, es decir, se realiza una asignación de red que es dirigir la señal a otras neuronas (sólo una señal por cada neurona), esto con el fin de reducir su funcionamiento a las neuronas básicas, es decir, las entradas se realizan en un formato vector, que al recibir dicha señal en su totalidad la red lo procesa y genera una salida con dicha información, que a su vez se convierte en entrada para la neurona receptora, y una vez que se haya extendido por toda la red, se produce un vector de salida. (Isasi Viñuela & Galvan León, 2004)

Ecuación de una salida de la red para dos capas:

$$\vec{S} = F(F(\vec{X} * W_1) * W_2) \quad (1)$$

Donde:

- W_1 y W_2 son los pesos de la primera y segunda capa, también son las matrices de conexión responsables de que el modelo aprenda.

- F es la función de activación y es idéntica para todas las neuronas.

- \vec{X} representa al vector de entrada a la red.

- \vec{S} hace referencia al vector de salida que produce la red.

Para comprender de mejor forma la ecuación de salida, se toma el siguiente ejemplo.

Supongamos que el vector de entrada es $\vec{x} = \{2, 1, -4\}$, con una función de activación $F(x) = 2x$ y los pesos,

$$W_1 = \begin{pmatrix} 2 & 1 & 3 \\ 3 & 2 & 2 \\ 1 & 2 & 1 \end{pmatrix} \text{ y } W_2 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$$

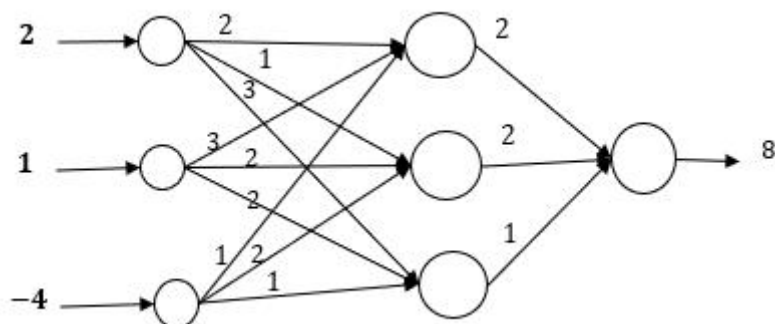
Entonces, de acuerdo a la ecuación 1, se tiene el producto del vector de entrada con el peso W_1 .

$$\vec{a} = \vec{x} * W_1 = (2, 1, -4) * \begin{pmatrix} 2 & 1 & 3 \\ 3 & 2 & 2 \\ 1 & 2 & 1 \end{pmatrix} = (3 \quad -4 \quad 4)$$

$$\vec{S} = F(F(\vec{a}) * W_2) = F((6 \quad -8 \quad 8) * \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}) = F((4)) = (8)$$

$$S = 8$$

En el ejemplo anterior se puede observar que la función de activación es lineal aplicada en dos capas de neuronas, y se puede representar gráficamente.

Figura 7*Red de dos capas*

La Figura 7 muestra el proceso del ejemplo anterior y su comportamiento del proceso que se realiza para llegar a la solución, la función de activación lineal es idéntica en ambas capas de la neurona.

El ejemplo anterior es para mostrar la aplicación del modelo de una red neuronal con dos capas y con función de activación lineal, pero en problemas complejos no solo se toma en cuenta la función, también es importante el momento en el cual para de aprender una neurona, para esto las neuronas en la red tienen un umbral asociado a estas, y en el caso del perceptrón multicapa este umbral es más como una conexión a la neurona, en el cual Isasi Viñuela y Galvan León (2004), indican que dicho umbral es igual a 1 y es constante.

Modelo del perceptrón multicapa

En un perceptrón multicapa, red con muchas capacidades, la salida de la señal de una neurona tiene sus cálculos por muchas funciones, además conocer las entradas, coeficientes de peso, métodos de agregación de entradas y características de la neurona permite que la señal de salida sea inequívoca en cualquier momento además de asumir que el proceso aparece de manera instantánea, una suposición teórica que no pasa con las neuronas

biológicas, ya que incluso en ensayos electrónicos toma tiempo para obtener una correcta señal de salida y adecuada. (Russell, 2018)

El perceptrón multicapa mediante la propagación hacia adelante define una relación entre la variable de entrada y salida, donde la neurona procesa la información y pasa a la siguiente capa.

Entonces, sea “C” el número de capas, “ n_c ” las neuronas en la capa “c” ($c=1, 2, \dots, C$), $W^c = w_{ij}^c$ la matriz de pesos asociados de la capa “c” a la capa “c+1” donde la conexión del peso va de la neurona i a la neurona j, $U^c = u_i^c$ el vector de umbrales y a_i^c la información de entrada a cada capa (de la neurona i a la neurona j) conocida como activación de la neurona.

La información de entrada $X = (x_1, x_2, \dots, x_{n1})$ en la primera capa o capa de entrada es igual a la activación a_i^c , cuando pasa a la siguiente capa se comporta como una primera salida.

El proceso en conjunto con la función de activación y el umbral se da en la capa oculta y es de la siguiente forma:

- **Activación en la capa oculta.** Para ello sea “f” una función de activación, entonces,

$$a_i^c = f \left(\sum_{j=1}^{n_{c-1}} w_{ji}^{c-1} a_j^{c-1} + u_i^c \right)$$

con $i = 1, 2, \dots, n_c$ y $c \neq 1 = \{2, 3, \dots, c\}$

Cabe aclarar que a_j^{c-1} corresponde a la activación en la capa c-1.

- **Activación en la capa de salida.**

$$y_i = a_i^c = f\left(\sum_{j=1}^{n_{c-1}} w_{ji}^{c-1} a_j^{c-1} + u_i^c\right)$$

con $i = 1, 2, \dots, n_c$

Donde el vector de salida es $Y = (y_1, y_2, \dots, y_{n_c})$ y C hace referencia a la última capa de salida.

El perceptrón multicapa, usa las funciones de activación sigmoideal y tangente hiperbólica con más frecuencia, esto al ser funciones continuas y crecientes, con intervalos que varían de $[0, 1]$ y $[-1, 1]$ respectivamente, y sus expresiones matemáticas son:

- Función sigmoideal

$$f(x) = \frac{1}{1 + e^{-x}}$$

- Función tangente hiperbólica

$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$$

Es importante tomar en cuenta que no existe una regla o método para determinar el número óptimo de capas ocultas o neuronas para la solución de un problema, si se tiene que obtener respuestas deseables es normal recurrir al error que comete o la comparación con valores de prueba, de esto tomar la decisión de aumentar o disminuir el número de neuronas y capas hasta lograr lo deseado, el proceso tomará tiempo por lo que la paciencia es la clave en la práctica.

2.1.5. Serie de tiempo

Para poder hacer mención sobre la serie de tiempo, es fundamental entender sobre los procesos estocásticos además de una variable aleatoria.

2.1.5.1. *Variable aleatoria*

Una variable aleatoria es una función que a cada suceso elemental de un espacio muestral se le asigna un elemento de la recta real. Cáceres Hernández (2006), menciona que, al analizar una sola característica sobre el resultado de un experimento, es conveniente realizar una transformación a dicho resultado en números reales, donde una variable aleatoria es una función y es la responsable de realizar dicha transformación en el cual se le asigna valores de la recta real a cada resultado del experimento.

Por ejemplo, sea el experimento de lanzar un dado, donde la variable aleatoria será representada por X , entonces esta variable representa el resultado del lanzamiento, es decir, un dado tiene seis caras y v.a. es una función, entonces $X(1) = 1, X(2) = 2, \dots, X(6) = 6$, donde el espacio muestral (S), que muestra los posibles resultados del experimento, es $S = \{1, 2, 3, 4, 5, 6\}$; además cada posible resultado tiene $P(X = x) = \frac{1}{6}$ de probabilidad de ser el resultado final.

Existen dos tipos de variables aleatorias y son:

✓ **Variable aleatoria discreta**

Una variable aleatoria discreta refiere a valores numerables o finitas, es decir que son contables, por ejemplo, el número de coches que van al garaje durante una semana, número de familias en un distrito, etc. (Maté Jimenez, 1995), en conclusión, una v.a. discreta es o son número enteros asignados a los resultados de un experimento, considerando el cero, en cuanto a los valores negativos podría darse por las pérdidas por ejemplo en el juego de apuestas.

Variable aleatoria continua

Maté Jimenez (1995) menciona que una v.a. continua puede tomar valores de la recta real, por ello no es posible conocer el valor exacto que toma la variable, porque la medición de su valor se clasifica en un intervalo, por ejemplo, medir el tiempo y el resultado podría ser de 72 segundos, de lo que podemos afirmar es que este tiempo es real no observable y se encuentra en un intervalo $[71.5, 72.5]$, esto hace la diferencia de una v.a. discreta. En estadística dichas magnitudes medidas se puede representar en histogramas de frecuencias relativas que, de acuerdo al autor, al tomar más observaciones empezara a tender a una curva suave en su comportamiento cuyo nombre recibe como una función de densidad.

Proceso estocástico

El proceso estocástico viene a ser una familia de variables aleatorias ordenadas en el tiempo. Mauricio (2007), menciona que un proceso estocástico viene a ser una secuencia de variables aleatorias, con un orden y equidistantes en el tiempo, que pueden ser procesos univariantes o escalar como también multivariante o vectorial, por lo cual, muestra una representación matemática de procesos estocásticos univariantes:

$$\dots, Y_{-1}, Y_0, Y_1, Y_2, \dots; (Y_t; t = 0, \pm 1, \pm 2, \dots)$$

Donde Y_t es una variable aleatoria en un tiempo t .

Mientras que, si $\mathbf{Y}_t \equiv [Y_{t1}, Y_{t2}, \dots, Y_{tM}]'$ con $M \geq 2$, sigue siendo una variable aleatoria vectorial, esto con el fin de mostrar o hacer referencia a un proceso estocástico multivariante, esto por el hecho de que se trabaja con más variables.

Series de tiempo

De acuerdo a Villavicencio (2010), una serie temporal son datos u observaciones en secuencia tomados en determinados momentos en el tiempo, con orden cronológico y

espaciados de manera uniforme, de esta forma los datos son dependientes entre si usualmente cuyo objetivo es realizar predicciones.

Serie temporal univariante, conocida como serie escalar, donde se observa una sola variable.

Mauricio (2007), muestra una representación matemática de series temporales univariantes:

$$y_1, y_2, \dots, y_n; (y_t)_{t=1}^n; (y_t: t = 1, \dots, n)$$

O también puede representarse en un vector columna, $\mathbf{Y} \equiv [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]'$ de orden n.

Donde y_t son las observaciones, “n” el tamaño de las observaciones y t toma valores de “1” hasta “n”, esto aplica para ambos tipos de series temporales, con la diferencia de que las multivariantes van en vectores columnas, para ser específico en matrices.

En una serie temporal se puede observar secuencias de observaciones en un orden cronológico, además de su variable o variables llevando así a las series univariantes o escalar, por contar con una sola variable dependiente siendo el tiempo como una variable independiente, y series multivariantes o vectorial, por tener dos o más variables.

Por esto se puede inferir que una serie temporal es un proceso estocástico, pero no viceversa como se muestra.

$$\dots, Y_{-1}, Y_0, Y_1, Y_2, \dots, Y_N, Y_{N+1}, \dots (\text{Proceso estocástico})$$

$$y_1, y_2, \dots, y_N (\text{Serie temporal})$$

Es importante recordar el objetivo de una serie temporal, consiste en elaborar un modelo estadístico que se ajuste adecuadamente a los datos y con ello describir la evolución de la serie, realizar la predicción y contrastar con evidencia teórica.

2.1.5.2. Características de las series temporales

Los datos de series temporales en su análisis toman tres características o componentes que deben cumplir, las cuales son:

1. Tendencia. Esta característica está relacionada con el cambio a largo plazo de la media, en el cual los datos muestran un comportamiento de manera creciente o decreciente.

La tendencia se puede observar gráficamente, pero existe una prueba que contrasta si la serie tiene o no tendencia.

Test Mann – Kendall, estadístico de contraste para probar la tendencia en la serie.

Teniendo,

$$S = \sum_{i=1}^{n-1} \sum_{j=k+1}^n \text{sign}(X_j - X_i)$$

Con **sign**($X_j - X_i$), función signo que indica la relación que hay entre X_j y X_i , además $j > i$ para evitar comparar elementos repetidos ni inversamente.

$$\text{sign}(X_j - X_i) = \begin{cases} +1 & \text{si } X_j - X_i > 0 \\ 0 & \text{si } X_j - X_i = 0 \\ -1 & \text{si } X_j - X_i < 0 \end{cases}$$

El test K-M se basa en lo siguiente,

$$z = \begin{cases} \frac{(S - 1)}{se} & \text{si } S > 0 \\ 0 & \text{si } S = 0 \\ \frac{(S + 1)}{se} & \text{si } S < 0 \end{cases}$$

Donde $se = \sqrt{var}$, ($var = \frac{1}{18} [n(n-1)(2n+5) - \sum f_t(f_t-1)(2f_t+5)]$)

(desviación estándar) y f_t representa al número de tiempos.

2. Estacionalidad. Esta característica se puede explicar como la presencia de periodicidad que hay en el tiempo, de manera semestral, mensual, trimestral, etc., un comportamiento repetitivo o similar en los periodos de tiempo.

Para encontrar la presencia de estacionalidad en una serie, se puede utilizar el estadístico de prueba Kruskal – Wallis, cuyo estadístico de contraste:

$$K = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2}$$

Donde n_i representa al número de observaciones en el grupo i, mientras que r_{ij} hace referencia al rango de la observación j en el grupo i, mientras que N es el total de observaciones de la serie.

$$\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i}, \bar{r} = \frac{(N-1)}{2} \text{ es el promedio de } r_{ij}.$$

3. Heterocedasticidad. En algunas bibliografías son nombradas como aleatorio o variabilidad haciendo referencia al comportamiento variable de los datos en ciertos momentos del tiempo.

Para encontrar la variabilidad que la serie puede presentar, se utiliza la prueba de Lagrange Multiplier (LM test), generalmente usado para modelos econométricos, dicha prueba es útil para ver si la serie requiere de un modelo volátil por presentar heterocedasticidad, su prueba de contraste está definida como:

$$LM = n * R^2$$

$$R^2 = 1 - \frac{SSR}{SST}$$

Con, número de las observaciones (n) y coeficiente de determinación del modelo auxiliar (R^2), suma de cuadrados de los residuos (SSR), suma total de cuadrados (SST), y

LM sigue una distribución χ^2 (chi-cuadrado) y los grados de libertad son iguales al número de retardos.

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2 \text{ y } SSR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Donde, \hat{Y}_i es el valor ajustado y \bar{Y} el promedio.

Por esto, una serie temporal se puede expresar como producto o suma de sus componentes:

$$X_t = T_t + E_t + A_t + \varepsilon_t \text{ (modelo aditivo)}$$

$$X_t = T_t * E_t * A_t * \varepsilon_t \text{ (modelo multiplicativo, } \varepsilon_t \text{ fluctúa entorno a 1)}$$

T_t representa a la tendencia, E_t a la estacionalidad, A_t heterocedasticidad y ε_t representa al componente aleatorio. (Villavicencio, 2010)

2.1.6. Metodología Box Jenkins

Miranda Chinlli (2021) indica que la metodología de Box Jenkins fue descrita en 1970 por los estadísticos George Edward Pelham Box y Gwilym Meirion Jenkins, con el objetivo de facilitar el trabajo a los analistas al momento de crear un modelo para una serie temporal.

Esta metodología tiene un uso amplio, su capacidad de manejar diferentes series, estacionarios o no estacionarios, hacen de su popularidad reconocida, además de su implementación en programas computacionales y de cómo se manejan los datos en el proceso, que va desde la verificación de la serie, identificación de un modelo tentativo, su estimación, diagnóstico y por último la predicción completa. (Ríos, 2008)

Esta metodología plantea cuatro pasos:

1. **Identificación:** En este paso se busca establecer los valores apropiados para p , d , q , P , D , Q . Se examina la serie temporal y se busca los patrones estacionales, tendencias y los posibles cambios en la varianza (estacionariedad) mediante pruebas estadísticas para ver si requieren diferenciación o algún tipo de transformación.
2. **Estimación:** Se debe estimar los parámetros incluidos en el modelo mediante técnicas de estimación como máxima verosimilitud o mínimos cuadrados. En este paso también es donde se identifica los órdenes del modelo que mejor se ajusten.
3. **Validación:** Al seleccionar el modelo, se debe verificar si este se ajusta bien a los datos, verificando la bondad de ajuste mediante el análisis de residuos viendo si cumplen con los supuestos mediante los test correspondientes y gráficos, la validación cruzada donde se divide los datos en dos grupos, de entrenamiento y prueba para evaluar el rendimiento del modelo.
4. **Pronóstico o predicción:** Por último, se realiza el pronóstico de la serie, establecer intervalos de confianza para medir la incertidumbre en las predicciones.

Esta metodología puede requerir de varias iteraciones hasta obtener el mejor modelo que satisfaga o se ajuste mejor a los datos y así capture adecuadamente las características de la serie temporal y proporcione pronósticos precisos.

Facilita la comprensión de la dinámica de la serie, además de permitir realizar predicciones.

2.1.6.1. Estacionariedad

En el análisis de una serie temporal el objetivo es encontrar un modelo que describa de mejor forma el comportamiento de los datos y así realizar la predicción, en términos

simples, se trata de acomodar a los modelos existentes según el comportamiento de la serie con los órdenes correctos con el propósito de encontrar la exactitud o aproximarse reduciendo el error de predicción, por ello se debe tener en cuenta si la serie es:

➤ **Estacionaria.** – La serie muestra dicha característica si su estabilidad es mostrada a lo largo del tiempo, media y varianza constantes.

Sea Y_t , es estacionario si la secuencia $Y_{t1}, Y_{t2}, \dots, Y_{tn}$ son semejantes a la secuencia $Y_{t1+h}, Y_{t2+h}, \dots, Y_{tn+h}$, con h tomando valores positivos mayores que 0.

Cuyas propiedades son:

$$\text{Media } E(Y_t) = E(Y_{t+h}) = \mu$$

$$\text{Varianza } V(Y_t) = V(Y_{t+h}) = E(Y_t^2) - (E(Y_t))^2 = \sigma^2$$

$$\text{Covarianza } \gamma_h = E[(Y_t - \mu)(Y_{t+h} - \mu)]$$

En conclusión, para que una serie sea estacionaria, su media, varianza y covarianza, en rezagos diferentes o en cualquier momento son iguales. (Villavicencio, 2010)

➤ **No estacionaria.** – la serie muestra variabilidad cambiante en el tiempo, es decir que, su media, varianza y covarianza son cambiantes a través del tiempo, esto conlleva a realizar procesos correspondientes para encontrar un modelo. (Ríos, 2008)

Sea Y_t , es no estacionario si las propiedades estadísticas de la secuencia $Y_{t1}, Y_{t2}, \dots, Y_{tn}$ son diferentes a la secuencia $Y_{t1+h}, Y_{t2+h}, \dots, Y_{tn+h}$, con h tomando valores positivos mayores que 0.

Mediante gráficos es posible observar su estacionariedad o no, es importante dicha característica en el análisis de series temporales debido a su valor al momento de tomar decisiones para encontrar el mejor modelo para su debida validación y predicción. No siempre es observable en la gráfica, si se trata de muchos datos, por ello, se puede contrastar

con la prueba de Dickey – Fuller si en verdad los datos muestran un comportamiento estacionario o requiere de una diferenciación o en otros casos de una transformación de ser necesario, y de acuerdo a ello poder deducir el tipo de modelo a aplicar en la serie.

2.1.6.2. Transformación Box Cox

La transformación Box Cox nace con la colaboración de los estadísticos George Box y Sir David Roxbee Cox en 1964. El supuesto de la normalidad es muy importante en la estadística, si los datos no cumplen con dicho supuesto se transforma variables dependientes que no siguen una distribución normal a seguir la normalidad.

De la Guía Gonzales (s.f.), menciona sobre la transformación de variables, esto debido a que la estadística encuentra datos que no cumplen con las características requeridas para su análisis, de esta forma requiriendo transformación a una o más variables, que en notación matemática vendría a ser $y = f(x)$, con el objetivo de tratar que los datos sigan una distribución normal.

La fórmula propuesta por Box Cox es:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0 \\ \log(y) & \text{si } \lambda = 0 \end{cases} \quad (2)$$

De acuerdo al sitio web Estadísticas Como hacerlo (s.f.), mencionan que la exponente lambda (λ), varia entre -5 a 5, pero cuyo valor óptimo varía de acuerdo a los datos que hacen su aproximación más cercana a una curva de la distribución normal.

Se debe tomar en cuenta que, si lambda toma el valor de 1, no hay transformación, es decir los datos permanecen igual, y cuando lambda se aproxima a cero la transformación hace que la varianza se estabilice.

2.1.6.2.1. Transformación inversa

Al final de la predicción, para realizar la comparación con la data de prueba y ver respecto a los valores reales, es necesario realizar una transformación regresando a la data original, entonces la ecuación es,

$$y = \begin{cases} ((y(\lambda) * \lambda) + 1)^{1/\lambda}, & \text{si } \lambda \neq 0 \\ e^{y(\lambda)} & \text{si } \lambda = 0 \end{cases}$$

Dicha ecuación servirá para realizar la transformación de los datos transformados, así como del pronóstico a su escala original.

2.1.6.3. Autocorrelación

Una vez comprobada que la serie es estacionaria, sigue el estudio de la correlación entre una observación y las observaciones previas, si existe correlación alta (autocorrelación) puede ser causado por un comportamiento lineal en el tiempo, y esto puede sugerir una idea sobre el tipo de modelo adecuado a utilizar, la forma visual que muestra es mediante correlogramas, un gráfico que muestra la correlación entre observaciones separadas por intervalos de tiempo (q) conocidos como “lags”, para su cálculo parte las observaciones en 2 grupos $\{Y_1, Y_2, \dots, Y_{t-q}\}$ y $\{Y_{1+q}, Y_{2+q}, \dots, Y_t\}$. (Ríos, 2008)

Villavicencio (s.f.) menciona que en ocasiones sucede que los valores de una variable no muestran ser independientes entre sí, es más, dependen de un valor anterior.

Hay dos maneras de ver las dependencias de las variables.

Función de autocorrelación simple (ACF)

Mide la correlación entre dos variables, por h periodos separados.

$$\rho_t = \text{corr}(Y_t, Y_{t-h}) = \frac{\text{cov}(Y_t, Y_{t-h})}{\sqrt{\text{Var}(Y_t)}\sqrt{\text{Var}(Y_{t-h})}}$$

Propiedades:

- $\rho_0 = 1$
- $-1 \leq \rho_t \leq 1$
- $\rho_t = \rho_{-t}$ (simetría)

Función de autocorrelación parcial (PACF)

Además de medir la correlación entre dos variables que se encuentran separadas por h periodos, en esta correlación no se considera la dependencia generada por retardos intermedios.

$$\pi_t = \text{corr}(Y_t, Y_{t-h} / Y_{t-1}, Y_{t-2})$$

$$\pi_t = \frac{\text{cov}(Y_t - \hat{Y}_t, Y_{t-h} - \hat{Y}_{t-h})}{\sqrt{\text{Var}(Y_t - \hat{Y}_t)} \sqrt{\text{Var}(Y_{t-h} - \hat{Y}_{t-h})}}$$

Auto covarianza

Sea Y_t un proceso estocástico, entonces la función de “Auto covarianza” relaciona valores a diferentes instancias. (Ríos, 2008)

$$\gamma_{(t,t+h)} = \text{Cov}(Y_t, Y_{t+h}) = E[(Y_t - E(Y_t))(Y_{t+h} - E(Y_{t+h}))]$$

Donde:

- $E(Y_t) \equiv \mu_Y$ (media)
- $\sigma_Y^2 \equiv \text{Var}(Y_t) \equiv E[(Y_t - \mu_Y)^2]$ (varianza)

Prueba de Ljung-Box

La prueba es útil para probar, de manera conjunta, que los coeficientes de autocorrelación son igual a cero, lo que indica que son independientes, en la cual trabaja con las siguientes hipótesis:

H_0 : Las autocorrelaciones son independientes.

H_a : Las autocorrelaciones no son independientes.

Esta se puede encontrar mediante la formula definida como

$$LB = n(n+2) \sum_{h=1}^m \left(\frac{\hat{\rho}_h^2}{n-h} \right) \sim \chi_m^2$$

Donde

- n representa al tamaño de la muestra.
- m es la longitud del rezago.
- h representa al número de rezagos.

Dicha prueba sigue una función chi-cuadrada, por ello es importante la tabla de dicha función, donde si el valor calculado es mayor al valor crítico de la tabla con un nivel de significancia asignada, se procede a rechazar la hipótesis nula, llegando a afirmar que por lo menos algunos de ellos muestran ser diferente de cero. (Villavicencio, 2010)

2.1.7. Modelo de serie de tiempo

2.1.7.1. Ruido blanco

Sea $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$, donde ε_t es una secuencia de variables idéntica e independientemente distribuidas con media 0 y varianza σ_ε^2 , si ε_t sigue una distribución normal, se representa como $\varepsilon_t \sim NIID(0, \sigma_\varepsilon^2)$, y se denomina como un proceso de ruido blanco “**normal o gaussiano**”. (Mauricio, 2007)

2.1.7.2. Paseo aleatorio

Conocido también como caminata aleatoria o camino aleatorio, esto del inglés “Random walk”, que de acuerdo al sitio web Britannica (2024), definen que en la teoría de probabilidades, un paseo aleatorio viene a ser un proceso para poder encontrar la ubicación probable de un punto que se encuentra en movimiento en alguna dirección, además de mencionar que siguen ejemplos de un proceso de Markov, esto porque el comportamiento futuro es independiente de su comportamiento pasado, mientras que Villavicencio (2010) menciona que un paseo aleatorio es un proceso estocástico, cuya primera diferencia viene a ser un ruido blanco ($\nabla Y_t = \varepsilon_t$), entonces podemos decir que un paseo aleatorio es el comportamiento impredecible en el tiempo de la serie.

Mauricio (2007) menciona que un paseo aleatorio es un proceso estocástico no estacionario, cuando es:

$$Y_t = \mu + Y_{t-1} + \varepsilon_t \text{ para } t = 0, \pm 1, \pm 2, \dots$$

En muchos casos μ vale cero, y $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$.

Si tenemos un modelo AR (1) no estacionario, donde un paseo aleatorio es $\nabla Y_t = \mu + \varepsilon_t$ que viene de $\nabla Y_t \equiv Y_t - Y_{t-1}$, tomando en cuenta que el proceso estocástico es univariante, se define a Y_t como un paseo aleatorio cuando la diferencia regular de orden uno es estacionario. (Mauricio, 2007)

2.1.8. Modelo ARIMA

El modelo ARIMA (Autoregressive Integrated Moving Average) es una herramienta poderosa y útil para el análisis de series temporales y la predicción de valores futuros que se puede aplicar en diferentes áreas como las finanzas para el análisis y pronóstico de los

precios ya sea de las acciones, tipos de cambio, precio de productos básicos y otros indicadores financieros, en salud para realizar el análisis y predicción de enfermedades, tasas de contagio, morbilidad, mortalidad y otros indicadores asociados al ámbito de la salud, y de esta forma a muchas otras áreas más donde se requiera este tipo de análisis y una predicción.

El modelo ARIMA es una combinación de 3 elementos principales, el componente autorregresivo (AR), captura las dependencias de los valores pasados, componente de mediana móvil (MA), captura la influencia de los errores pasados, y el componente de diferenciación o integración (I) para convertir la serie no estacionaria en estacionaria y eliminar las tendencias y patrones no estacionarios.

Modelo autorregresivo (AR)

En este modelo se representa las observaciones pasadas de una variable como sumas ponderadas, con “p” número de rezagos además de representar el orden de este modelo.

Modelo AR (p)

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

Con $\varepsilon_t \sim N(0, \sigma^2)$, μ y ϕ_1 parámetros, $|\phi_1| < 1$ por condición de estacionariedad.

Modelo AR (2)

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

➤ **Esperanza**

$$E(Y_t) = E(c) + E(\phi_1 Y_{t-1}) + E(\phi_2 Y_{t-2}) + E(\varepsilon_t)$$

$$E(Y_t) = c + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2})$$

$$E(Y_t) = E(Y_{t-1}) = \mu \text{ “estacionariedad de la serie”}$$

$$E(Y_t) = \mu = c + \phi_1 \mu + \phi_2 \mu$$

$$\mu = \frac{c}{1 - \phi_1 - \phi_2}$$

Despejando la constante c tenemos, $c = \mu - \phi_1\mu - \phi_2\mu$, para luego reemplazar en la ecuación del modelo AR (2).

$$Y_t = \mu - \phi_1\mu - \phi_2\mu + \phi_1Y_{t-1} + \phi_2Y_{t-2} + \varepsilon_t$$

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \varepsilon_t$$

De esta manera se obtiene el modelo AR (2) en desviaciones a su media.

$$\tilde{Y}_t = \phi_1\tilde{Y}_{t-1} + \phi_2\tilde{Y}_{t-2} + \varepsilon_t$$

➤ Varianza

$$Var(Y_t) = \gamma_0 = Cov(Y_t, Y_t) = E(Y_t - \mu)^2 = E(\tilde{Y}_t)^2$$

$$\gamma_0 = E(\phi_1\tilde{Y}_{t-1} + \phi_2\tilde{Y}_{t-2} + \varepsilon_t)^2$$

$$\gamma_0 = E(\phi_1^2\tilde{Y}_{t-1}^2 + \phi_2^2\tilde{Y}_{t-2}^2 + \varepsilon_t^2 + 2\phi_1\tilde{Y}_{t-1} * \phi_2\tilde{Y}_{t-2} + 2\phi_1\tilde{Y}_{t-1} * \varepsilon_t + 2\phi_2\tilde{Y}_{t-2} * \varepsilon_t)$$

$$\gamma_0 = \phi_1^2E(\tilde{Y}_{t-1}^2) + \phi_2^2E(\tilde{Y}_{t-2}^2) + E(\varepsilon_t^2) + 2\phi_1\phi_2E(\tilde{Y}_{t-1} * \tilde{Y}_{t-2})$$

$$\gamma_0 = \phi_1^2\gamma_0 + \phi_2^2\gamma_0 + 2\phi_1\phi_2\gamma_1 + \sigma^2$$

$$\gamma_0 = \frac{2\phi_1\phi_2\gamma_1 + \sigma^2}{1 - \phi_1^2 - \phi_2^2} \quad (1)$$

Tomando en cuenta el resultado de la ecuación (2) lo reemplazamos en la ecuación de la varianza (1).

$$\gamma_0 = \frac{2\phi_1\phi_2\left(\frac{\phi_1\gamma_0}{1 - \phi_2}\right) + \sigma^2}{1 - \phi_1^2 - \phi_2^2}$$

$$\gamma_0 = \frac{\left(\frac{2\phi_1^2\phi_2\gamma_0 + \sigma^2(1 - \phi_2)}{1 - \phi_2}\right)}{1 - \phi_1^2 - \phi_2^2}$$

$$\gamma_0 = \frac{(2\phi_1^2 \phi_2 \gamma_0 + \sigma^2(1 - \phi_2))}{(1 - \phi_1^2 - \phi_2^2)(1 - \phi_2)}$$

$$\gamma_0 = \frac{\sigma^2(1 - \phi_2)}{1 - \phi_1^2 - \phi_2^2 - \phi_2 + \phi_2\phi_1^2 + \phi_2^3 - 2\phi_1^2 \phi_2}$$

Y de esta manera se obtiene la varianza del modelo AR (2) y es,

$$\gamma_0 = \frac{\sigma^2(1 - \phi_2)}{1 - \phi_1^2 - \phi_2^2 - \phi_2 + \phi_2^3 - \phi_1^2 \phi_2} = \frac{\sigma^2(1 - \phi_2)}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]}$$

➤ **Autocovarianza**

$$\mathbf{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-1}) = \boldsymbol{\gamma}_1 = \mathbf{E}((\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-1} - \boldsymbol{\mu})) = \mathbf{E}(\widetilde{\mathbf{Y}}_t * \widetilde{\mathbf{Y}}_{t-1})$$

$$\gamma_1 = E((\phi_1 \widetilde{Y}_{t-1} + \phi_2 \widetilde{Y}_{t-2} + \varepsilon_t) \widetilde{Y}_{t-1})$$

$$\gamma_1 = \phi_1 E(\widetilde{Y}_{t-1}^2) + \phi_2 E(\widetilde{Y}_{t-2} \widetilde{Y}_{t-1})$$

$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1$$

$$\gamma_1 = \frac{\phi_1 \gamma_0}{1 - \phi_2} \quad (2)$$

$$\mathbf{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-2}) = \boldsymbol{\gamma}_2 = \mathbf{E}((\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-2} - \boldsymbol{\mu})) = \mathbf{E}(\widetilde{\mathbf{Y}}_t * \widetilde{\mathbf{Y}}_{t-2})$$

$$\gamma_2 = E((\phi_1 \widetilde{Y}_{t-1} + \phi_2 \widetilde{Y}_{t-2} + \varepsilon_t) \widetilde{Y}_{t-2})$$

$$\gamma_2 = \phi_1 E(\widetilde{Y}_{t-1} * \widetilde{Y}_{t-2}) + \phi_2 E(\widetilde{Y}_{t-2}^2)$$

$$\gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0$$

$$\mathbf{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-3}) = \boldsymbol{\gamma}_3 = \mathbf{E}((\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-3} - \boldsymbol{\mu})) = \mathbf{E}(\widetilde{\mathbf{Y}}_t * \widetilde{\mathbf{Y}}_{t-3})$$

$$\gamma_3 = E((\phi_1 \widetilde{Y}_{t-1} + \phi_2 \widetilde{Y}_{t-2} + \varepsilon_t) \widetilde{Y}_{t-3})$$

$$\gamma_3 = \phi_1 E(\widetilde{Y}_{t-1} * \widetilde{Y}_{t-3}) + \phi_2 E(\widetilde{Y}_{t-2} * \widetilde{Y}_{t-3})$$

$$\gamma_3 = \phi_1 \gamma_2 + \phi_2 \gamma_1$$

Generalizando se puede deducir que,

$$\gamma_i = \phi_1 \gamma_{i-1} + \phi_2 \gamma_{i-2}, \text{ con } i \geq 2$$

➤ **Autocorrelación**

Al dividir por γ_0 se obtiene:

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\frac{\phi_1 \gamma_0}{1 - \phi_2}}{\gamma_0} = \frac{\phi_1}{1 - \phi_2}$$

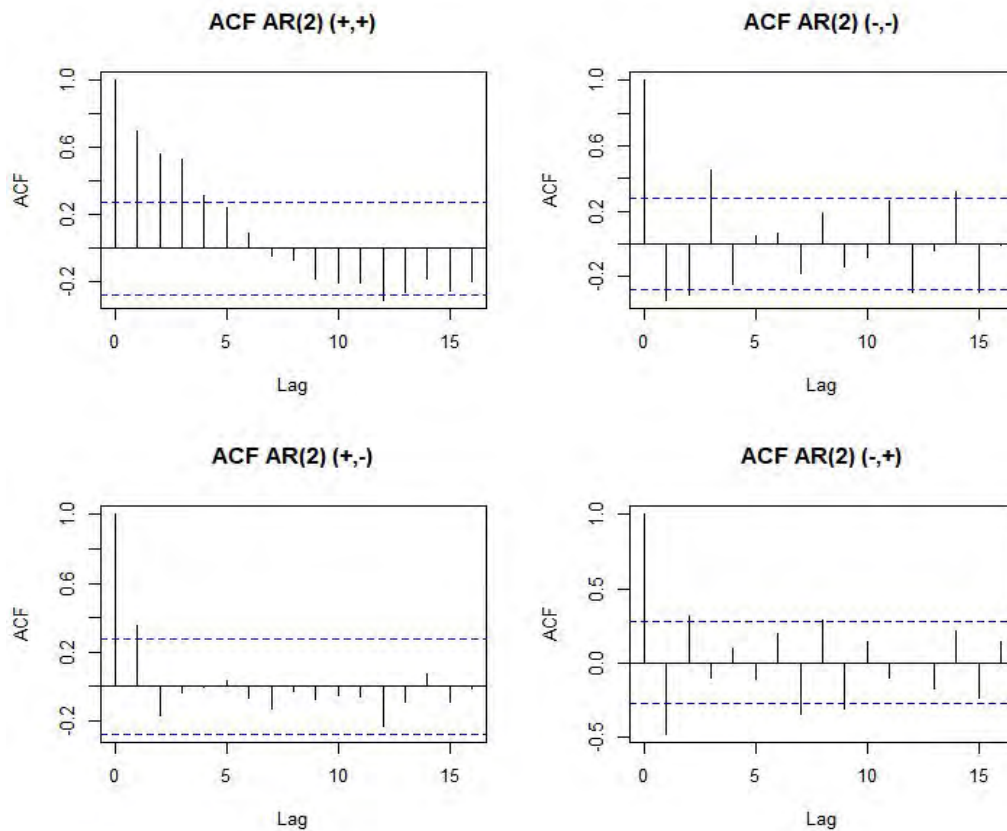
$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1 \gamma_1 + \phi_2 \gamma_0}{\gamma_0} = \phi_1 \rho_1 + \phi_2$$

Generalizando, se concluye que:

$$\rho_n = \frac{\gamma_n}{\gamma_0} = \phi_1 \rho_{n-1} + \phi_2 \rho_{n-2}$$

Figura 8

Gráficos de autocorrelación de modelo AR (2)



En la Figura 8 se muestra cuatro gráficos de autocorrelación, de acuerdo al signo que toman los parámetros, para el modelo autorregresivo de orden 2. Cuando los signos son positivos las barras o líneas apuntan hacia arriba, cuando los signos son diferentes intercalan las líneas, uno arriba y el otro hacia abajo.

Modelo de media móvil (MA)

En este modelo, se representa una suma ponderada de los errores actuales y anteriores, q representa el número de rezagos del error además representa el orden del modelo.

Modelo MA (q)

$$Y_t = c + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Con $\varepsilon_t \sim N(0, \sigma^2)$, μ y θ_1 parámetros, $|\theta_1| < 1$ por condición de invertibilidad.

Modelo MA (2)

$$Y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$

➤ **Esperanza**

$$E(Y_t) = E(\mu) + E(\theta_1 \varepsilon_{t-1}) + E(\theta_2 \varepsilon_{t-2}) + E(\varepsilon_t)$$

$$E(Y_t) = \mu + \theta_1 E(\varepsilon_{t-1}) + \theta_2 E(\varepsilon_{t-2}) + E(\varepsilon_t)$$

$$E(Y_t) = \mu$$

➤ **Varianza**

$$Var(Y_t) = \gamma_0 = E(Y_t - \mu)^2$$

$$\gamma_0 = E(\mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t - \mu)^2$$

$$\gamma_0 = \sigma^2 + \theta_1^2 \sigma^2 + \theta_2^2 \sigma^2$$

➤ **Autocovarianza**

$$Cov(Y_t, Y_{t-1}) = \gamma_1 = E((Y_t - \mu)(Y_{t-1} - \mu))$$

$$\gamma_1 = E((\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t)(\theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + \varepsilon_{t-1}))$$

$$\gamma_1 = \sigma^2(\theta_1 + \theta_1 \theta_2)$$

$$\text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-2}) = \boldsymbol{\gamma}_2 = \mathbf{E}((\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-2} - \boldsymbol{\mu}))$$

$$\gamma_2 = E((\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t)(\theta_1 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-4} + \varepsilon_{t-2}))$$

$$\gamma_2 = \sigma^2 \theta_2$$

$$\text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-3}) = \boldsymbol{\gamma}_3 = \mathbf{E}((\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-3} - \boldsymbol{\mu}))$$

$$\gamma_3 = E((\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t)(\theta_1 \varepsilon_{t-4} + \theta_2 \varepsilon_{t-5} + \varepsilon_{t-3}))$$

$$\gamma_3 = 0$$

Por ello, se concluye que, para cualquier rezago mayor o igual a 3, la covarianza de dicho orden es 0.

➤ Autocorrelación

Al dividir por γ_0 se obtiene:

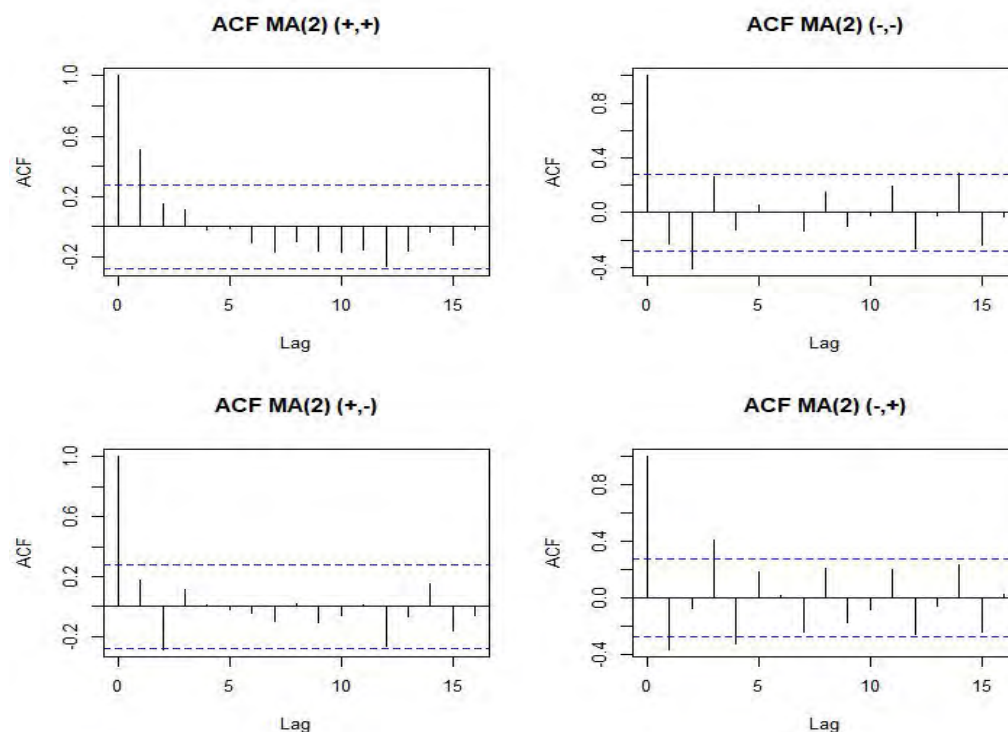
$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\sigma^2(\theta_1 + \theta_1 \theta_2)}{\sigma^2(1 + \theta_1^2 + \theta_2^2)} = \frac{\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_1 = \frac{\gamma_2}{\gamma_0} = \frac{\sigma^2 \theta_2}{\sigma^2(1 + \theta_1^2 + \theta_2^2)} = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

Figura 9

Gráfico de autocorrelación del modelo MA (2)



En la Figura 9 se muestra cuatro gráficos de autocorrelación, considerando las diferentes combinaciones de signos que pueden tomar sus parámetros, para el modelo de medias móviles de orden 2, en los gráfico se muestra el comportamiento del coeficiente de autocorrelación para distintos retardos, y las líneas de confianza que delimitan el rango en el que las autocorrelaciones pueden considerarse estadísticamente no significativas.

Modelo ARMA

Este modelo es una representación mixta de los modelos AR(p) y MA(q), en el que se observa en función de los rezagos actuales y pasados además de las observaciones pasadas.

Modelo ARMA (p, q)

$$Y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Donde $\varepsilon_t \sim N(0, \sigma^2)$.

Modelo ARMA (1, 1)

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

➤ **Esperanza**

$$E(Y_t) = E(c + \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1})$$

$$E(Y_t) = E(Y_{t-1}) = \mu \text{ “estacionariedad de la serie”}$$

$$E(Y_t) = c + \phi_1 \mu$$

$$E(Y_t) = \mu = \frac{c}{1 - \phi_1}$$

➤ **Varianza**

Para ello reemplazaremos la esperanza obtenida en la ecuación del modelo con el objetivo de centrar el proceso:

$$Y_t = \mu - \mu\phi_1 + \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\tilde{Y}_t = \phi_1 \tilde{Y}_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Partiendo desde aquí,

$$\text{Var}(Y_t) = \gamma_0 = \text{Cov}(Y_t, Y_t) = E(Y_t - \mu)^2 = E(\tilde{Y}_t)^2$$

$$\gamma_0 = E(\phi_1 \tilde{Y}_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1})^2$$

$$\gamma_0 = \phi_1^2 \gamma_0 + \sigma^2 + \theta_1^2 \sigma^2$$

$$\gamma_0 = \frac{\sigma^2(1 + \theta_1^2)}{1 - \phi_1^2}$$

➤ **Autocovarianza**

$$\gamma_1 = \text{Cov}(Y_t, Y_{t-1}) = E((Y_t - \mu)(Y_{t-1} - \mu)) = E(\tilde{Y}_t * \tilde{Y}_{t-1})$$

$$\gamma_1 = E((\phi_1 \tilde{Y}_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1})(\tilde{Y}_{t-1}))$$

$$\gamma_1 = \phi_1 \gamma_0$$

$$\gamma_2 = \text{Cov}(Y_t, Y_{t-2}) = E((Y_t - \mu)(Y_{t-2} - \mu)) = E(\tilde{Y}_t * \tilde{Y}_{t-2})$$

$$\gamma_2 = E((\phi_1 \tilde{Y}_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1})(\tilde{Y}_{t-2}))$$

$$\gamma_2 = \phi_1 \gamma_1$$

Por lo tanto, la autocovarianza para un orden mayor a 2, en el modelo ARMA (1,1) seguirá la forma,

$$\gamma_k = \phi_1 \gamma_{k-1}$$

➤ Autocorrelación

Al dividir por γ_0 se obtiene:

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\phi_1 \gamma_0}{\gamma_0} = \phi_1$$

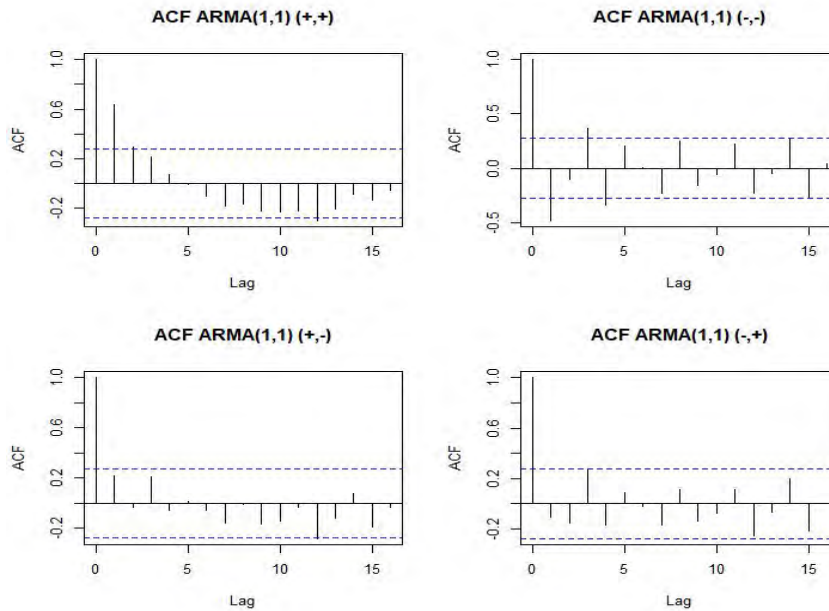
$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1 \gamma_1}{\gamma_0} = \phi_1 \rho_1$$

Por lo tanto, en general sigue la forma,

$$\rho_k = \phi_1 \rho_{k-1}$$

Figura 10

Gráficos de autocorrelación, ARMA (1,1)



En la Figura 10 se muestra cuatro gráficos de autocorrelación, de acuerdo al signo que toman los parámetros, para el modelo ARMA (1,1).

En la vida real, hay series que no son estacionarios, y esta necesidad conlleva a encontrar un modelo que puede manejar este tipo de series, realizando una diferenciación logra que su media y varianza sean constantes a lo largo del tiempo, y el modelo que puede hacer esto es el ARIMA.

Miranda (2021), indica que una serie temporal $\{X_t\}$ sigue un modelo ARIMA, si es un proceso ARMA (p, q) y es estacionario, decimos que $\{X_t\}$ es un proceso ARIMA (p, d, q).

Modelo general ARIMA (p, d, q)

$$\phi(B)(1 - B)^d Y_t = c + \theta(B)\varepsilon_t$$

Donde:

- Y_t es el valor de la serie temporal en el instante t.

- c es una constante o el término de la media de la serie temporal.
- p es el orden del componente autorregresivo (AR), el cual representa la cantidad de términos autorregresivos incluidos en el modelo.
- ϕ es el coeficiente autorregresivo que representa la influencia de los valores pasados.
- d es el orden de diferenciación y representa la cantidad de veces que se aplicó la diferenciación para convertir la serie en estacionaria.
- q es el orden del componente de medias móviles (MA) e indica la cantidad de términos del promedio móvil que están incluidos en el modelo.
- θ es el coeficiente de la media móvil (MA), el cual indica la influencia de los errores pasados.
- ε_t conocido como ruido blanco o el error en el momento t .
- B es el operador de retardo.

Modelo ARIMA (1,1,2)

$$(1 - \phi_1 B)(1 - B)^1 Y_t = c + (1 + \theta_1 B + \theta_2 B^2) \varepsilon_t$$

$$Y_t = c + (1 - \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

El objetivo del modelo ARIMA es estimar los valores óptimos de los coeficientes, mediante la estimación de máxima verosimilitud o el método de mínimos cuadrados, de tal forma minimizar el error de predicción y obtener un modelo que se ajuste mejor a los datos de la observación.

Es importante tener en cuenta que primero se realiza la estimación de los coeficientes para luego poder utilizar el modelo para realizar las predicciones futuras y analizar la serie temporal.

2.1.9. Modelo SARIMA

El modelo SARIMA es una extensión del modelo ARIMA, el proceso SARIMA es un proceso ARIMA estacional, lo que indica que, como anteriormente ya se mencionó que para que el proceso de una serie temporal $\{X_t\}$ sea un proceso ARIMA (p, d, q) debe ser estacionario y por ello eliminar la tendencia, en el modelo SARIMA es sugerible eliminar la tendencia aplicando la diferenciación (d) regular ante una serie con tendencia, este modelo normalmente trabaja con series no estacionarios pero es preferible trabajar con series estacionarios para evitar problemas futuros en el modelo.

Según Miranda (2021), el modelo SARIMA refleja la característica de variación estacional en series de tiempo.

Modelo SARIMA o ARIMA (o ARIMA estacional multiplicativo)

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^DY_t = c + \theta_1(B)\Theta_1(B^s)\varepsilon_t$$

Donde:

- Y_t es el valor de la serie temporal en el instante t.
- c es una constante o el término de la media de la serie temporal.
- p es el orden del componente autorregresivo (AR), el cual representa la cantidad de términos autorregresivos incluidos en el modelo.
- P es el componente autorregresivo (AR) en la parte estacional.
- ϕ es el coeficiente autorregresivo que representa la influencia de los valores pasados.
- φ es el coeficiente autorregresivo en la parte estacional.
- d es el orden de diferenciación y representa la cantidad de veces que se aplicó la diferenciación para convertir la serie en estacionaria.

- D es el orden de diferenciación en la parte estacional.
- q es el orden del componente de medias móviles (MA) e indica la cantidad de términos del promedio móvil que están incluidos en el modelo.
- Q es el orden de (MA) en la parte estacional.
- θ es el coeficiente de la media móvil (MA), el cual indica la influencia de los errores pasados.
- Θ es el coeficiente de (MA) en la parte estacional.
- ε_t conocido como ruido blanco o el error en el momento t .
- B es el operador de retardo.
- S es la frecuencia estacional (número de periodos en un ciclo estacional).

Modelo SARIMA (1,1,1) (1,1,1)

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)^1(1 - B^{12})^1 Y_t = c + (1 + \theta_1 B)(1 + \Theta_1 B^{12}) \varepsilon_t$$

$$Y_t = c + (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + (1 + \phi_1)Y_{t-12} - (1 + \phi_1 + \Phi_1 + \phi_1 \Phi_1)Y_{t-13} + (\phi_1 + \phi_1 \Phi_1)Y_{t-14} - \Phi_1 Y_{t-24} + (\Phi_1 + \phi_1 \Phi_1)Y_{t-25} - \phi_1 \Phi_1 Y_{t-26} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \Theta_1 \varepsilon_{t-12} + \theta_1 \Theta_1 \varepsilon_{t-13}$$

El modelo SARIMA combina las componentes estacionales y las no estacionales para modelar adecuadamente la serie temporal que muestra patrones estacionales y tendencias.

2.1.9.1. Estimación de parámetros

Máxima verosimilitud

Miranda Chinlli (2021), menciona que la estimación mediante máxima verosimilitud, los valores de los parámetros hacen que los valores de los datos observados sean los más probables, es decir, los valores maximizan la función de verosimilitud.

Se conoce que el ruido blanco sigue una distribución normal, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, cuya función es,

$$f(\varepsilon_t) = (2\pi\sigma_\varepsilon^2)^{-1/2} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}\right) \quad (i)$$

para esto nos apoyaremos del modelo $AR(1)$.

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

Ahora consideraremos centrar la variable a su media.

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t$$

De donde podemos encontrar la ecuación en función del ruido blanco.

$$\varepsilon_t = (Y_t - \mu) - \phi(Y_{t-1} - \mu) \quad (ii)$$

Ahora, de la ecuación (i), obtenemos.

$$L(\phi, \mu, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-n/2} \exp\left(-\frac{\sum_{t=1}^n (\varepsilon_t^2)}{2\sigma_\varepsilon^2}\right)$$

Sea $S(\phi, \mu) = \sum_{t=1}^n (\varepsilon_t^2)$. Aplicando el logaritmo a la función, se tiene,

$$\ell(\phi, \mu, \sigma_\varepsilon^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} S(\phi, \mu)$$

de esto aplicando la función score (U) e igualando a cero, se tiene,

$$\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\phi}, \hat{\mu})}{n}$$

Para valores datos a los parámetros (ϕ, μ) se puede maximizar de manera analítica con respecto a σ_ε^2 , esto porque para dichos parámetros, no hay una solución analítica directa, esto porque conllevan derivadas parciales que implican ecuaciones no lineales, por esto se calcula ε_t y se ajusta los parámetros hasta maximizar ℓ .

2.1.9.2. Criterio de evaluación para la predicción

Para elegir el modelo adecuado, primero se tiene que observar detalladamente la serie, mediante gráficos y pruebas resaltando las características, y de ser necesario se recurre a una transformación de la serie cuando no siguen una distribución normal o muestran sesgos, si los datos son estacionarios o no hace mucha diferencia al momento de identificar el modelo, así como los patrones variables, los gráficos ACF y PACF son buenos ayudantes al momento de poder identificar el orden tentativo.

El **criterio de información de Akaike (AIC)**, es útil para identificar el mejor modelo que se ajuste a los datos, permite también la búsqueda del orden en el modelo, mediante minimización.

$$AIC = -2\ln(\ell) + 2r$$

Donde:

- ℓ es la máxima verosimilitud.
- r , representa al número de parámetros, incluyendo la constante, del modelo.

Para verificar la capacidad predictiva del modelo, hay indicadores que cuantifican la diferencia que existe con respecto a la serie, y de acuerdo a ello poder escoger el modelo para el pronóstico.

Error absoluto medio (MAE)

Mide la precisión del pronóstico promediando los errores cometidos por el pronóstico.

$$MAE = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n}$$

Con “n” el número de observaciones.

Raíz cuadrática media del error (RMSE)

Ríos (2008), indica que mide la dispersión que tienen los datos en la variable, elevando los errores al cuadrado donde penaliza fuertemente los errores grandes.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t^s - Y_t^a)^2}$$

Y_t^s , valor real de la serie.

Y_t^a , valor predicho.

n , es el número de periodos.

Promedio del error porcentual Absoluto (MAPE)

Conocido también como “error porcentual absoluto medio”, nace a partir de la necesidad de ver los errores en forma porcentual, su cálculo viene de encontrar los errores absolutos en cada periodo, dividido entre el valor observado para dicho periodo y

promediando los errores porcentuales absolutos, dicha técnica es útil cuando la serie tiene valores grandes. (Lazo Chuquihuayta, 2019)

$$MAPE = \frac{1}{T} \left(\sum_{t=1}^T \frac{|Y_t^s - Y_t^a|}{Y_t^a} \right) \times 100$$

Y_t^s , valor real de la serie.

Y_t^a , valor predicho.

T, es el número de periodos.

Algo a tomar en cuenta es, tener un MAPE menor no siempre hace que se generará una predicción con un RMSE menor o viceversa. (Ríos, 2008)

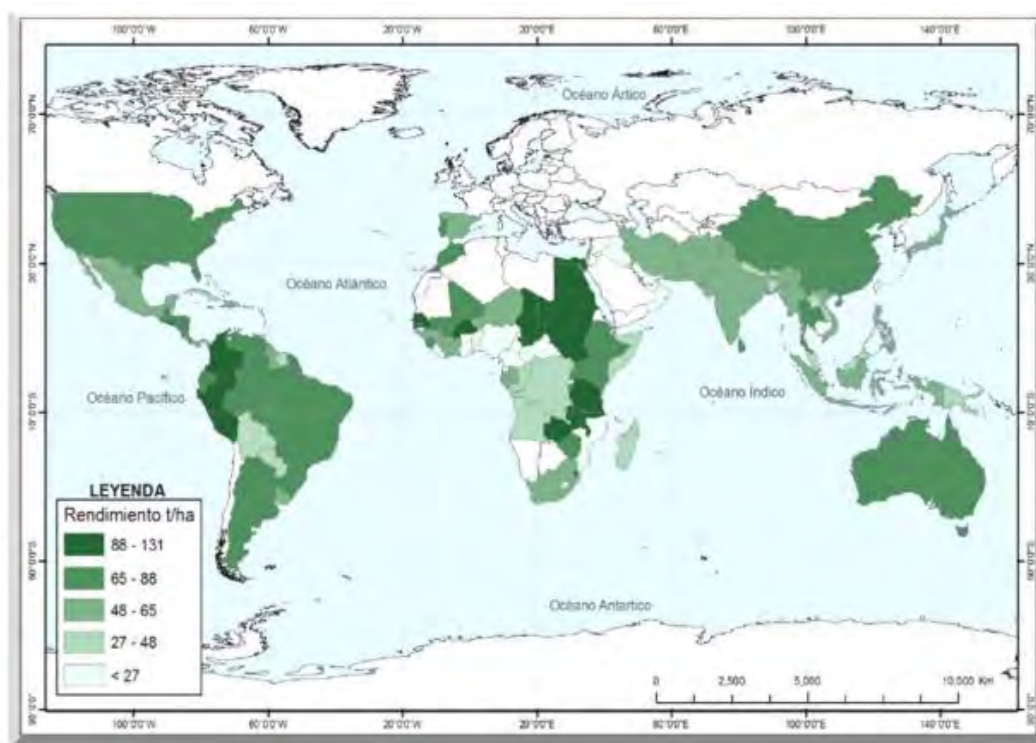
2.2. Marco conceptual

2.2.1. Caña de azúcar (*Saccharum officinarum* L.)

De acuerdo al Comité Nacional para el Desarrollo Sustentable de la Caña de Azúcar (2015), la caña de azúcar es una gramínea con origen en Nueva Guinea (Sureste de Asia y la India occidental, 327 a.c.), según la historia 647 d.c. fue introducido en Egipto y más tarde en España (755 d.c.), el cultivo de la caña de azúcar se extendió por todo el mundo, y en américa por Cristóbal Colón en su viaje.

Figura 11

Países productores de caña de azúcar en el mundo



Fuente: FAOSTAT, 2011, citado por InfoAgro.

La Figura 11 muestra los países productores de azúcar y caña de azúcar en el mundo mediante un mapa, resaltando con un color verde e indicando mientras más oscuro es la representación del país que produce más en el mundo, dicha información pertenece al año 2011.

Es importante mencionar que este producto almacena hidratos de carbono en forma de sacarosa, y toma importancia la parte del tallo, Ballat et al. (2014) mencionan que el tallo está conformado por una serie de entrenudos, dentro y fuera de la tierra, el tallo es cilíndrica o ligeramente aplanada con un tamaño promedio de 2 a 4 metros y diámetro de 3 a 5 cm, la parte superior del tallo contiene poca sacarosa lo que conlleva a su eliminación, pero para que el producto crezca es importante que el clima sea cálido o una zona sin heladas además

de mucha agua sin descuidar el suelo o terreno de producción, debe estar fértil. El ciclo de este producto, desde la plantación hasta la cosecha varía entre los 10 a 24 meses.

El sitio web SofOS (2020), menciona sobre la importancia de la predicción sobre la producción de caña de azúcar, debido al riesgo que conlleva por el precio en el mercado y el clima, así como otros factores impredecibles; productores enfrentan desafíos en el día a día tomando control en el proceso como la limpieza de malezas y plagas además del riego controlado, cuyas decisiones no siempre son acertadas. Por ello menciona la existencia de dos estrategias de predicción, estrategias sin observaciones (análisis de datos históricos sobre el rendimiento y clima para futuros rendimientos del cultivo) y estrategias con observaciones (proceso productivo ya iniciado mediante muestras en el que se pesa la cantidad de tallos de un determinado espacio, imágenes satelitales o drones).

La predicción mediante muestras puede traer más beneficios que el método de drones, a pesar de la dificultad la información puede ser vital para tomar decisiones al aplicar la cantidad de agua necesaria, así como el control de plagas y malezas, además de poder decidir si adelantar o retrasar el corte, eliminar la producción si no hay remedio o salvar de alguna manera el cultivo, por ello la estadística es fundamental para la toma de decisiones.

2.2.2. Aptitud del uso del suelo para el cultivo de caña de azúcar

Es importante tomar en cuenta las condiciones ambientales adecuadas, como es la humedad, la temperatura, calidad del suelo o la tierra, etc., para el desarrollo de un cultivo, las practicas ganaderas o forestales pueden contribuir de buena manera aportando beneficios al suelo en aumentar sus nutrientes y el cultivo tenga lo suficiente para su crecimiento y producción.

Miranda, Marza, Calle et al. (2021), en su investigación citan a conservación internacional, donde muestran un cálculo de la cantidad de nutrientes necesarios para un rendimiento de 90 t ha, con un 85 kg de nitrógeno, 180 kg de fósforo y 180 kg de potasio se puede lograr dicha cantidad; los suelos sufren limitaciones en cuanto al fósforo así como precipitaciones excesivas, en el estudio realizado a base de las muestras lograron captar que la caña de azúcar absorbe 100 kg de nitrógeno por ha, mientras que clasificaron al terreno como marginalmente apto para el cultivo de caña de azúcar por su bajo contenido de fósforo, de acuerdo a otras investigaciones debe ser superior a 40 kg a 100 kg por hectárea para que sea óptimo para la siembra; recalca que un suelo ácido limita el desarrollo de la caña de azúcar.

La calidad del suelo para cultivar caña de azúcar requiere de estudios previos, si bien es cierto hay estudios realizados en diferentes partes del mundo, pero no todos coinciden en cuanto a los nutrientes que requiere el cultivo para su desarrollo, esto debido por la geografía, el clima y la cantidad de nutrientes que el suelo contiene además de la temperatura, sin embargo, mientras que haya suficiente agua y el suelo cumpla ciertos parámetros, se puede cultivar el producto, más tener en cuenta que la cantidad de producción variará de acuerdo a su tratamiento o seguimiento en el proceso de desarrollo del cultivo.

2.2.3. Azúcar (*Sacarosa*)

El nombre azúcar es un término coloquial que se le dio a la sacarosa, nombre real del producto, que en química se puede explicar como un disacárido a base de glucosa y fructuosa obtenida de la caña de azúcar o remolacha azucarera, es bien conocido los tipos de azúcar: moreno, rubio y blanco, nombrados así por su apariencia en cuanto al color, este último (azúcar blanco) sometido a un proceso de purificación (centrifugación). Su historia

viene de un aproximado pasado de 5000 años con los árabes que en sus conquistas o invasiones introdujeron en diversos reinos como es Siria, Egipto y el norte de África, así como en reinos vecinos, cuyo refinamiento es el logro de químicos egipcios lo cual es la sacarosa, y más tarde convirtiendo un alimento conocido en todo el mundo, pero no hasta la edad media con su llegada a España. (Ministerio de Agricultura, Pesca y Alimentación, s.f.)

El azúcar no contiene proteínas, minerales y vitaminas, pero aun así es fundamental para el buen funcionamiento de diferentes órganos, una de sus funciones es producir energía que el cuerpo necesita, mientras que ciertas hormonas como la insulina (hormona producida por el páncreas cuya función es regular los niveles de glucosa en la sangre) mantienen estable el flujo de glucosa, si baja, la persona podría sufrir debilidad, temblor y hasta desmayos llegando a una situación crítica, por ello es bueno tener conocimiento que su consumo excesivo puede llevar a contraer enfermedades afectando a los dientes, la obesidad o diabetes.

Producción de azúcar en el mundo

El consumo de azúcar es invariable, ya sea mediante bebidas o alimentos que contienen dicho producto, conlleva a conocer los países con producción mayor en el mundo, y de acuerdo a MIDAGRI (2023), quien a su vez se basaron en la información de USDA (Departamento de Agricultura de los Estados Unidos), Brasil es el mayor productor de azúcar liderando desde hace varios años atrás con 42 millones de toneladas para el año 2023, y con más detalle se muestra en la siguiente tabla.

Tabla 1*Países con mayor producción de azúcar en el mundo*

Posición	País	2023	Porcentaje
1	Brasil	42010	22.4%
2	India	36000	19.2%
3	Unión Europea	15475	8.2%
4	China	10000	5.3%
5	Tailandia	11200	6.0%
6	Rusia	6336	3.4%
7	Estados Unidos	8369	4.5%
8	México	6254	3.3%
9	Pakistán	7110	3.8%
10	Australia	4400	2.3%
21	Perú	1300	0.7%
	Otros países	39427	21.0%
Producción total		187881	100%

Fuente: MIDAGRI (2023)

De la Tabla 1 se destacan Brasil e India, siendo los países con mayor producción en el mundo, con 22% y 19% respectivamente, además podemos encontrar a Perú en el puesto 21 con una producción de 1.3 millones de toneladas, a pesar de la cantidad, la ONU, indica que lidera en cuanto al rendimiento de caña de azúcar en el mundo.

Procesos de producción del azúcar

El azúcar o sacarosa se da del proceso de cristalización del jugo de la caña (en el jugo se encuentra el 20% de azúcares, 14% de fibra y 65% de agua), pero el proceso empieza por la **recepción** de la caña de azúcar recién cosechada o fresca para obtener más contenido de sacarosa, acto seguido se **tritura** la caña con el objetivo de extraer la sacarosa (agua caliente facilita el molido) del cual se obtiene el bagazo (residuos del tallo del cual se elabora el papel), la siguiente etapa es eliminar las impurezas (**clarificación**), luego se realiza el proceso de **evaporación** del agua (80% del jugo) y para ello es necesario hervir en el cual se utiliza el bagazo como leña generando calor que conlleva a su evaporación, el proceso de la

cristalización se realiza en los tachos de cocimiento, mediante presión y temperatura determinada, de donde se forman cristales, la **centrifugación** se realiza para separar los cristales de azúcar de la maleza (mediante fuerza centrífuga hace que la masa cocida suba por la pared exterior de la centrifugadora mientras que el azúcar es detenida mediante un filtro, por último se realiza el **secado** en el que se tamiza los cristales con el objetivo de separar los de tamaño adecuado para su envase. (Ballat et al, 2014)

2.3. Antecedentes

2.3.1. Antecedentes internacionales

Dip y Romero (2015), en la revista de investigación en modelos financieros, “Una comparación de redes neuronales y modelos ARCH-GARCH para predecir variaciones en el precio de acciones, aplicación a un caso de acciones de telefonía”, donde el objetivo de dicho trabajo fue contrastar la teoría de eficiencia de mercado demostrando la posibilidad de predicción del precio de la acción de Telecom Argentina S.A., mencionan que la búsqueda de modelos de predicción más eficientes y exactos para las variaciones se viene dando desde hace tiempo, en ese transcurso se determinó la existencia de dos modelos para la predicción de series temporales (modelo lineales y no lineales), con las redes neuronales perteneciendo al segundo grupo, donde las nuevas exigencias del mercado á marcado un cambio en la búsqueda del perfeccionamiento y sofisticación de herramientas utilizadas para la predicción, de esa forma llevando a una evolución en la predicción.

Miranda Chinlli (2021), en su tesis titulada “Modelización de series temporales modelos clásicos y SARIMA” cuyo objetivo general es presentar la metodología de Box Jenkins para modelos SARIMA, que ayude a pronosticar la temperatura de la ciudad de New York, utilizando datos históricos de series de tiempo, recopilados de la estación

meteorológica por el proyecto CRUTEM, menciona que debido al calentamiento global en nuestro planeta es necesario tomar precaución al momento de investigar las razones de este cambio climático, por esto es muy importante la predicción de la información meteorológica, de esta forma poder tomar decisiones y acciones para salvaguardar especies vegetales y animales a los cuales comprometen amenazas a su conservación. Concluye que la metodología de Box y Jenkins es eficaz para ajustar modelos de series temporales de forma sencilla en las predicciones, además de que estos modelos son útiles a corto plazo.

Ruiz, Hernández y Zulueta (2011), en su investigación “Análisis de series de tiempo en el pronóstico de la producción de caña de azúcar”, donde el objetivo es prever el comportamiento de su proceso productivo con el fin de planificar y optimizar el uso de los recursos técnicos, humanos y financieros, mencionan que la permanencia y competencia juegan un rol en la búsqueda e implementación de métodos para un pronóstico con una mayor precisión y de esta forma poder tomar las decisiones necesarias en la industria azucarera, de este modo, mediante la aplicación del modelo Box – Jenkins, lograron encontrar el modelo ARIMA(1,2,0) logrando una precisión en un 94%. Los datos contenían información desde 1949 a 2006 cuyo pronóstico fue para el intervalo de tiempo 2006 a 2007, cuyo resultado de la producción anual sería de 11,974 toneladas de azúcar, pero la producción real fue de 12,736 toneladas, a lo que justifican que la variación se dio por la disminución de la superficie sembrada en un 1.5%.

Isasi Viñuela y Galvan León (2004), en su libro “Redes neuronales artificiales, un enfoque práctico”, cuyo enfoque radica en el paradigma matemático para la resolución de problemas más que las relaciones entre las redes neuronales y sus analogías computacionales, mencionan que el objetivo de una red neuronal artificial es que las máquinas procesen

paralelamente con elementos neuronales, donde su comportamiento global “emule” de la mejor forma posible, además de que las neuronas tienen una función de transición de estado o de activación que mediante la recepción de información o señales (proveniente del exterior o neuronas a las que están conectadas) puede cambiar de estado, por ejemplo si $S=\{0, 1\}$, se denomina al conjunto de estados posibles de una neurona, entonces 0 vendría a ser estado inactivo mientras que el 1 sería es estado activo. Se toma en cuenta que S puede tomar más valores que en este caso sería una imagen de $n+1$ niveles de gris.

2.3.2. Antecedentes nacionales

Menacho (2013), en su investigación “Comparación de los métodos de series de tiempo y redes neuronales” cuyo objetivo es comparar los métodos estadísticos clásicos aplicados al análisis de series de tiempo con los modelos de redes neuronales artificiales multicapa back propagación, menciona que en las últimas décadas está en aumento el uso del modelo de redes neuronales artificiales (RNA), porque son auto adaptativos además de inferir resultados puede obtener soluciones aun si hay presencia de datos faltantes, su uso va abarcando desde problemas de clasificación hasta métodos alternativos como son los problemas de predicción porque no requieren muchos supuestos, por ello concluyen con la ventajosa respuesta hacia el modelo de redes neuronales multicapa con un resultado de predicción preciso, resaltando su capacidad de aprendizaje y su auto adaptabilidad con una recomendación de experimentar con otros tipos de redes neuronales.

Lazo Chuquihuayta (2019), en su tesis “Modelo de redes neuronales artificiales para el pronóstico del número de visitantes extranjeros a Machu Picchu en comparación con la metodología de Box y Jenkins”, cuyo objetivo general es comparar el modelo de redes neuronales artificiales para el pronóstico de número de visitantes extranjeros a Machu Picchu

2002 – 2018 con la metodología Box y Jenkins, menciona la búsqueda de un mejor modelo de predicción para obtener información sobre el comportamiento futuro que tendrán las visitas extranjeras a la ciudadela de Machu Picchu, de esta forma llega a la conclusión de la RNA frente al modelo de Box y Jenkins (ARIMA), captura de mejor forma el comportamiento de los datos y por ende su mejor precisión del pronóstico.

Puma Cárdenas (2020), en su tesis “Modelos Autorregresivos Integrado de Media Móvil y Red Neuronal Recurrente para predecir la demanda de medicamentos, hospital regional Cusco, 2013 – 2018”, donde el objetivo principal es analizar los modelos autorregresivos integrado de media móvil (ARIMA) o la red neuronal recurrente tipo de memoria a corto y largo plazo (RNNLSTM) para predecir adecuadamente la demanda de medicamentos en el hospital regional del Cusco, 2018, menciona que bajo la muestra de cinco tipos de medicamentos (cloruro de sodio, Ibuprofeno, Metamizol sódico, Paracetamol y Sevoflurano) y la aplicación de ambos modelos, llegó a la conclusión de que la RNNLSTM predice de mejor manera frente a un ARIMA, esto basándose en los resultados del MAPE con un bajo porcentaje de error cometido.

III HIPÓTESIS Y VARIABLES

3.1. Hipótesis

3.2. Hipótesis general

El modelo de la red neuronal artificial permite una mejor predicción de la producción azúcar en el Perú.

3.3. Hipótesis específicas

Los modelos SARIMA, de la metodología Box y Jenkins, proporcionan un mejor ajuste para la predicción de la producción de azúcar en el Perú.

Los modelos de red neuronal artificial, tipo perceptrón multicapa, proporcionan un mejor ajuste para la predicción de la producción de azúcar en el Perú.

El modelo SARIMA y la red neuronal artificial presentan diferencias en su capacidad de ajuste y predicción, siendo uno de ellos adecuado para realizar predicciones más precisas.

3.4. Identificación de variables e indicadores

Variable dependiente

Producción de azúcar

Variable independiente

Tiempo (1994 – 2023)

3.6. Operacionalización de variables

Tabla 2

Operacionalización de variables

Variables	Definición conceptual	Naturaleza	Escala
Dependiente: Producción de azúcar	El azúcar es un endulzante natural cristalizado hecho principalmente de sacarosa, que se obtiene de plantas como la caña de azúcar o la remolacha azucarera a través de procesos industriales. (MIDAGRI, 2023)	Cuantitativa	Razón
Independiente: Tiempo (1994 – 2023)	El tiempo tiene varias interpretaciones relacionadas a la cronología y según la RAE es una magnitud física que permite el orden de secuencias de los sucesos, diferenciando el pasado, presente y futuro, su unidad de medida en el sistema internacional es el segundo.	Cuantitativo	Razón

IV METODOLOGÍA

4.1. Ámbito de estudio

La presente investigación abarca a todos los departamentos productores de azúcar del Perú.

4.2. Tipos y nivel de investigación

4.2.1. Tipo de investigación

La presente investigación es aplicativa, debido a que se utilizan los conocimientos de la estadística con el fin de poder aplicar en el proceso de producción de azúcar a futuro, con un pronóstico adecuado que aproximará la cantidad de producción de azúcar, el cual influirá en los ingresos a las familias productoras y al país.

4.2.2. Nivel de investigación

La investigación es descriptivo, longitudinal (1994 – 2023) y predictivo porque se realiza la predicción para el año 2024.

4.3. Población de estudio

La población de estudio comprende la secuencia de la cantidad de producción total mensual azúcar, recolectada desde el mes de enero de 1994 hasta el mes de diciembre de 2023 lo cual se encuentra en la base de datos del Banco Central de Reserva del Perú (BCRP), donde están ordenados y equidistantes en diferentes momentos en el tiempo.

4.4. Técnicas de recolección de datos

Para el presente trabajo de investigación, los datos se obtuvieron de la página web del Banco Central de Reserva del Perú (BCRP), en la sección de estadísticas, en el cual el archivo está disponible en formato Excel, los cuales corresponden a la cantidad total de producción mensual de azúcar información de enero de 1994 hasta diciembre de 2023.

4.5. Técnicas de análisis e interpretación de la información

1. Procesamiento de los datos

Primeramente, se ordenó los datos conforme a los meses y años correspondientes (1994 a 2023), para posteriormente haciendo uso de la estadística descriptiva identificar el comportamiento de los datos mediante gráficos (histograma, gráfico de dispersión y el grafico de cajas) en donde se evaluó si hay presencia de datos faltantes y datos atípicos.

Después se describe las características presentadas respecto a la producción de azúcar (1994 – 2023) utilizando gráficos temporales, funciones de autocorrelación simple y parcial, también test para ver la estacionariedad, tendencia, estacionalidad y heterocedasticidad y tener una idea más clara de que método poder aplicar, también se usó las pruebas de normalidad y estacionariedad.

2. Software utilizado

Este proceso se realizó utilizando el programa Excel y el software libre R-Project (4.3.1) con un conjunto de paquetes complementarios del mismo para el análisis, cálculo y gráficos de la información.

4.6. Técnicas para demostrar la verdad o falsedad de las hipótesis planteadas

Para demostrar la veracidad o falsedad de las hipótesis plantadas, se utilizó la raíz del error cuadrático medio (RMSE) y el error porcentual absoluto medio (MAPE).

V RESULTADOS

5.1. Análisis, interpretación y discusión de resultados

5.1.1. Metodología Box Jenkins

La serie consta de 360 observaciones sobre la producción de azúcar en el Perú en tiempo mensual, 1994 - 2023, tomando el periodo enero a diciembre del año 2023 para ver la precisión del pronóstico realizado mediante la metodología Box Jenkins, y de acuerdo al análisis de dichos datos en la Figura 12 se puede observar el comportamiento de la serie en el tiempo.

Tabla 3

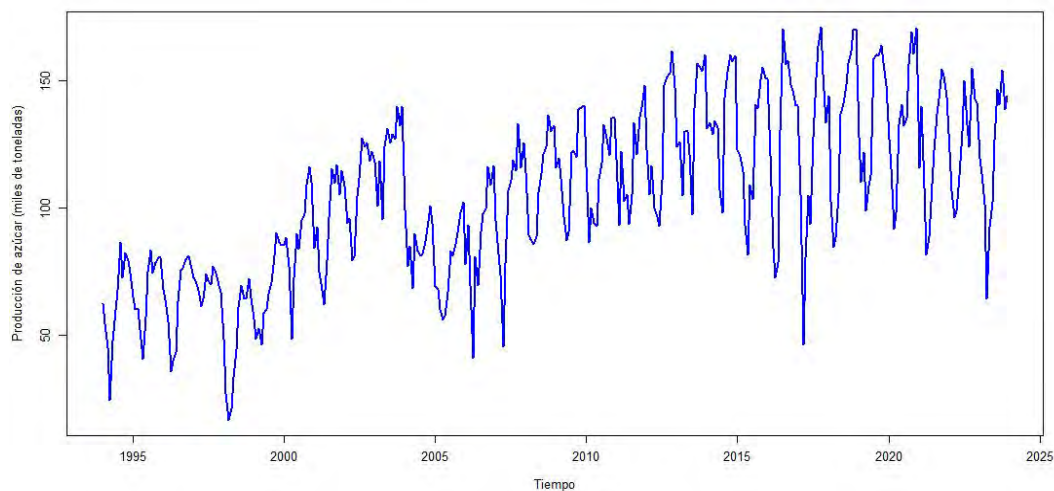
Cuadro resumen de estadísticos de la producción de azúcar en el Perú

Estadístico	Valor
Mínimo	16.81
1er cuartil	80.59
Mediana	105.23
Media	105.08
3rd cuartil	131.26
Máximo	170.94

En Tabla 2 se observa que el mínimo de la producción fue 16.81 miles de toneladas así también alcanzando los 170.94 miles de toneladas con un promedio de 105.08 miles de toneladas, además el 25% de la producción fue menor a 80.59 miles de toneladas, el 50% fue menor a 105.23 miles de toneladas y el 75% llegó hasta 131.26 miles de toneladas de producción de azúcar mensual en el transcurso de 1994 a 2023 en el Perú.

Figura 12

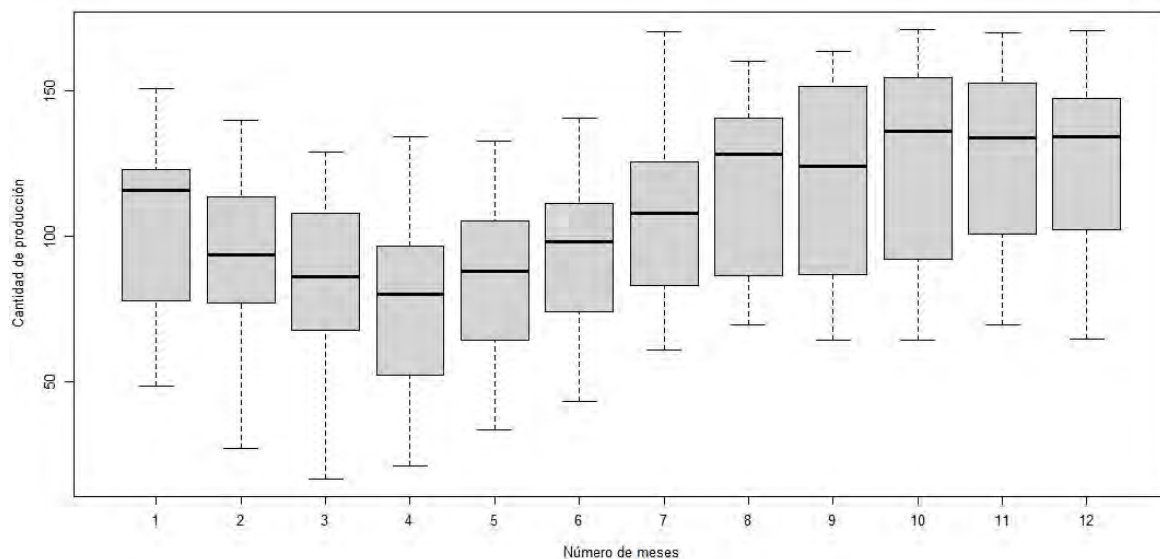
Comportamiento de la producción de azúcar en el tiempo (1994 – 2023)



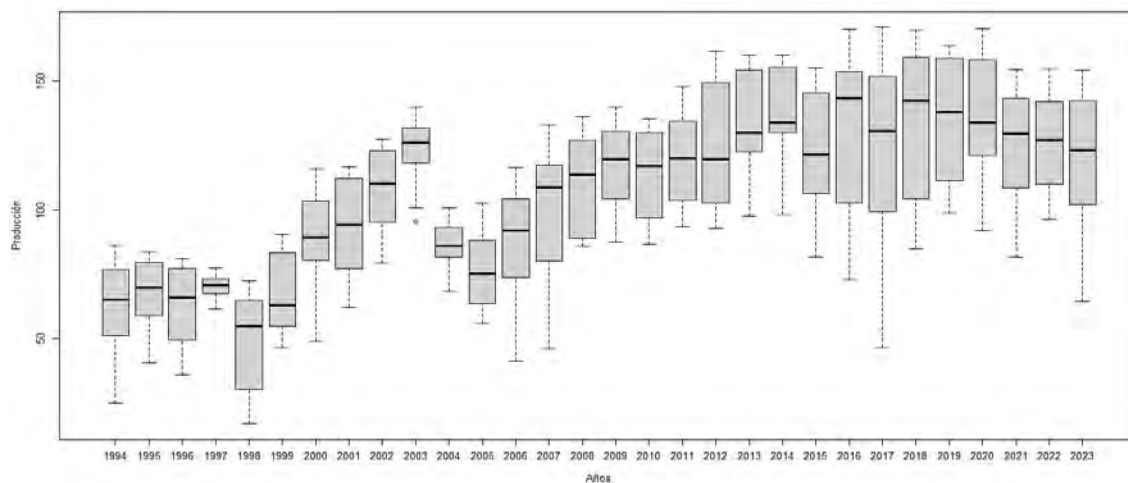
La Figura 12 de la producción de azúcar muestra una tendencia positiva realizando su continuo incremento, a pesar de ello la baja producción en ciertos meses y años muestran un patrón de incremento, esto relacionado al costo alto que llegan a tener cuando la producción baja, de esta manera mostrando poca variabilidad, así como una estacionalidad aparente. Por ello sugiere un modelo multiplicativo por ser lo adecuado, por sus componentes multiplicativos.

Figura 13

Comportamiento de los datos por mes en un diagrama de cajas

**Figura 14**

Comportamiento de los datos por años en un diagrama de cajas



El diagrama de cajas es útil al momento de detectar valores atípicos que podrían afectar al modelo, en la Figura 13 que muestra el comportamiento de los datos mensualmente, se puede observar que no hay datos atípicos además de mostrar una posible

estacionalidad, mientras que en la Figura 14 se puede observar que en el año 2003 hubo una producción muy baja, siendo de 95 miles de toneladas en el mes de abril, mientras que en dicho año la producción en promedio alcanzo los 123 miles de toneladas, en su mayoría por encima de 100 mil toneladas mensuales mostrando una diferencia menor considerando la cantidad.

Adecuación de la data para series temporales

Para poder realizar las predicciones correspondientes, es importante tener la certeza de contar con el modelo adecuado y los pronósticos adecuados, para ello se toma en consideración dividir la serie, tomando 348 datos como entrenamiento (1994-2022) y 12 datos para la prueba, este último corresponde a la serie anual del 2023 como prueba para su comparación, por ello es necesario hacer el análisis correspondiente con la data de entrenamiento.

De acuerdo al test de Lilliefors (versión modificada del test de Kolmogorov – Smirnov para probar el supuesto de normalidad cuando se desconoce los parámetros y deben ser estimados a partir de los datos), $D = 0.041992$ y $p\text{-value} = 0.1275$, los datos tienen una aproximación a una distribución normal siendo el valor P de probabilidad de contraste mayor al nivel de significancia (5%) no mostrando evidencia suficiente para rechazar la hipótesis nula, además que la diferencia máxima entre la distribución observada y normal tiene un aproximado de 0.042, indicando que la diferencia es baja y afirmando su aproximación a un comportamiento normal.

5.1.1.1. Estacionariedad

Hipótesis

H_0 : La serie tiene una raíz unitaria (no es estacionaria)

H_1 : La serie no tiene una raíz unitaria (es estacionaria)

Tabla 4

Test Dickey Fuller para probar la estacionariedad de la serie

Test Dickey - Fuller

P-Value 0.01

Bajo el test de Dickey – Fuller, tomando en cuenta el valor P de probabilidad de contraste es menor al 5% (nivel de significancia), como se observa en la Tabla 3, rechazar la hipótesis nula, sin evidencia estadística que lo contradiga, y afirmando que la serie de producción de azúcar es estacionaria.

5.1.1.2. Características que presenta la serie

➤ Tendencia

Hipótesis

H_0 : La serie no presenta tendencia

H_1 : La serie presenta tendencia

Tabla 5

Test Mann Kendall para probar la tendencia de la serie

Test Mann Kendall

P-Value 2.22×10^{-16}

Bajo el test de Mann Kendall, tomando en cuenta el valor P de probabilidad de contraste es menor al 5% (nivel de significancia), como se observa en la Tabla 4, rechazar la hipótesis nula al no haber evidencia estadística que lo contradiga, y afirmando que la serie de producción de azúcar presenta tendencia.

➤ **Estacionalidad**

Hipótesis

H_0 : La serie no presenta estacionalidad

H_1 : La serie presenta estacionalidad

Tabla 6

Test Kruskal - Wallis para probar la estacionalidad de la serie

Test Kruskal – Wallis	
P-Value	2.067 x 10 ⁻¹⁴

Bajo el test de Kruskal – Wallis, tomando en cuenta el valor P de probabilidad de contraste es menor al 5% (nivel de significancia), como se muestra en la Tabla 5, al no encontrar evidencia estadística que contradiga se rechaza la hipótesis nula afirmando que la serie de producción de azúcar presenta estacionalidad.

➤ **Heterocedasticidad**

Hipótesis

H_0 : La serie no presenta heterocedasticidad condicional en los datos (los errores no presentan patrones de variabilidad cambiante en el tiempo)

H_1 : La serie presenta heterocedasticidad condicional en los datos (los errores presentan patrones de variabilidad cambiante en el tiempo)

Tabla 7*Test Lagrange Multiplier para probar la variabilidad de la serie*

LM test	
P-Value	2.2×10^{-16}

Bajo el LM test, tomando en cuenta el valor P de probabilidad de contraste es menor al 5% (nivel de significancia), como se observa en la Tabla 6, podemos rechazar la hipótesis nula, al no tener evidencia estadística que contradiga, afirmando que la serie de producción de azúcar presenta variabilidad cambiante en el tiempo.

Las pruebas realizadas para conocer las características que presenta la serie, muestran que la serie de la producción de azúcar presenta estacionariedad, tendencia, estacionalidad y variabilidad cambiante con el tiempo.

Transformación Box Cox. Tomando el valor de lambda ($\lambda = 0.46$) en la ecuación 2, se realiza la transformación de la serie.

Para comprobar si hay presencia de heterocedasticidad, se recurre al test de Levene, cuya hipótesis nula indica que no hay heterocedasticidad (varianzas iguales), entonces se obtiene un P-valor de 0.997 que es mayor al 5% nivel de significancia, lo que indica que, al no encontrar evidencia estadística para rechazar la hipótesis nula, concluimos con que no hay presencia de heterocedasticidad en los datos transformados.

Representación gráfica de la función de autocorrelación simple y parcial

Figura 15

Representación gráfica de la autocorrelación simple

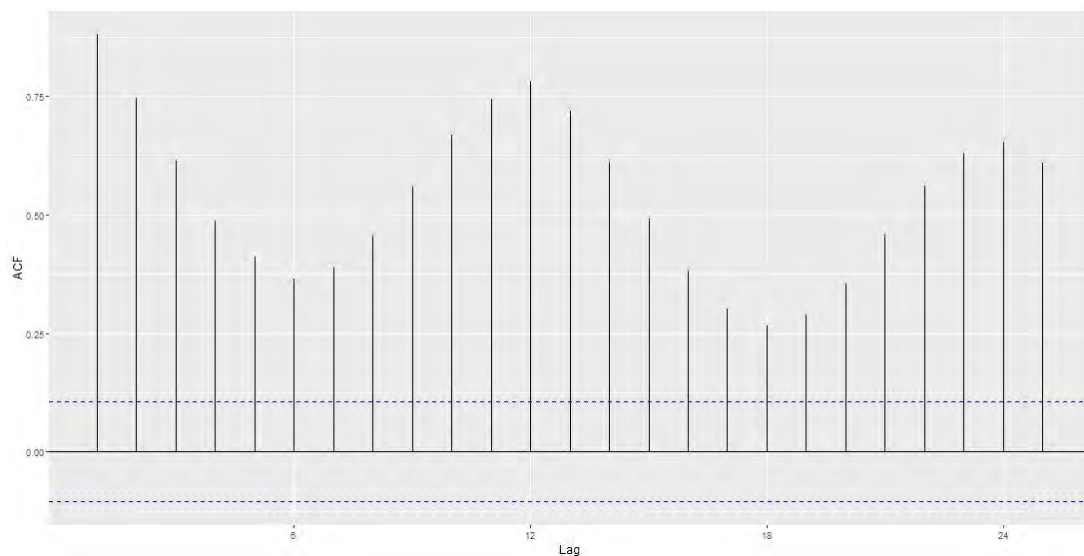
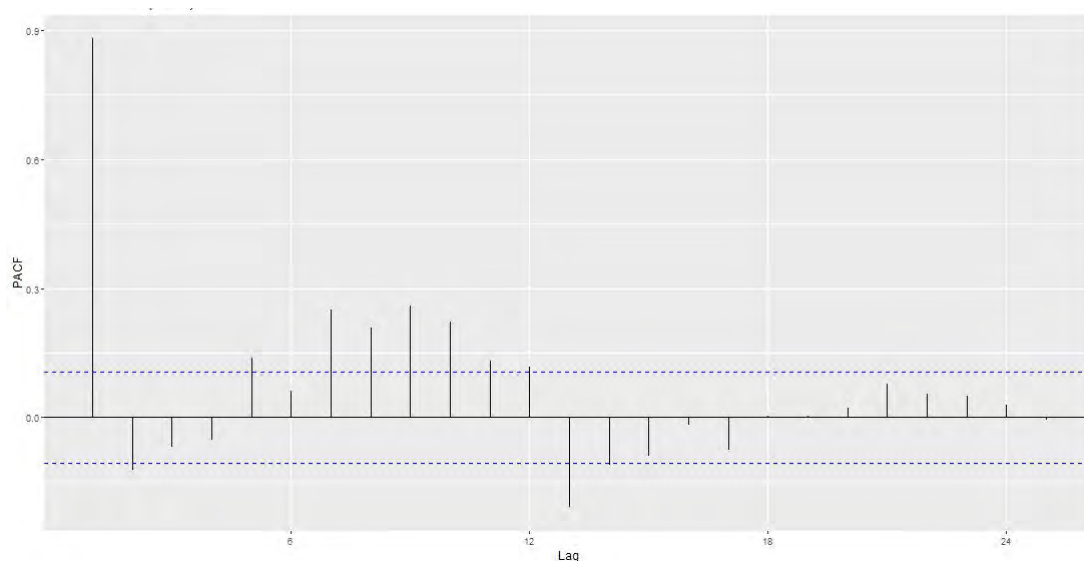


Figura 16

Representación gráfica de la autocorrelación parcial



En la Figura 15 de autocorrelación simple se puede observar una oscilación, esto se da debido a la presencia de estacionalidad que presenta la serie, además de mostrar que los datos son estacionarios; entonces la estacionalidad y la tendencia a corto plazo son los

causantes de dicha oscilación, mientras que en la Figura 16 de autocorrelación parcial es posible observar el orden que tomara la parte autorregresiva del modelo que puede ser 1, 2, 3 y hasta 4 mientras que el orden para las medias móviles pueden ser de orden 1, 2 y 3 posiblemente.

Una de las características que cumple la serie es la estacionalidad y por esto podemos deducir que el modelo a utilizar es SARIMA para su debida predicción.

Tabla 8

Comparación de modelos tentativos para la serie producción de azúcar

Modelo	Medidas de ajuste			Precisión del pronóstico			AIC
	RMSE	MAPE	MAE	RMSE	MAPE	MAE	
SARIMA(4,0,0)(0,1,1) ₁₂	1.070	5.371	0.806	0.964	4.781	0.817	1039.970
SARIMA(3,0,1)(0,1,1)₁₂	1.025	5.095	0.760	0.955	3.879	0.600	1023.550
SARIMA(3,0,2)(0,1,1) ₁₂	1.024	5.071	0.756	0.966	3.958	0.613	1024.830
SARIMA(3,0,3)(0,1,1) ₁₂	1.021	5.052	0.753	1.102	4.727	0.747	1024.910
SARIMA(4,0,3)(0,1,1) ₁₂	0.990	4.897	0.732	0.971	4.353	0.684	1021.000

Nota: MAPE (Error porcentual absoluto medio); RMSE (Raíz cuadrada media del error); AIC (Criterio de información de Akaike); MAE (Error absoluto medio).

En la Tabla 7 se puede observar cinco modelos a elegir, de acuerdo a los valores menores de AIC y los indicadores estadísticos (MAE, RMSE, MAPE), siendo el modelo SARIMA (4,0,3) (0,1,1)₁₂ quien se ajusta mejor a la serie, porque muestra indicadores estadísticos con valores menores que el resto, así como un menor AIC = 1021, sin embargo es el modelo SARIMA (3,0,1) (0,1,1)₁₂ quien muestra tener mejor ajuste en la predicción, mostrando indicadores menor al resto y un AIC = 1023.55 que comparado con el modelo anterior no varía mucho, lo que conlleva a concluir sobre el modelo con mejor ajuste de predicción en la metodología Box Jenkins.

SARIMA (3,0,1) (0,1,1)₁₂ con AIC = 1023.55

$$[1 - (\phi_1 B + \phi_2 B^2 + \phi_3 B^3)](1 - B^{12})Y_t = c + (1 + \theta_1 B)(1 + \Theta_1 B^{12})\varepsilon_t$$

$$Y_t(1 - B^{12} - \phi_1(B - B^{13}) - \phi_2(B^2 - B^{14}) - \phi_3(B^3 - B^{15})) = c + \varepsilon_t(1 + \Theta_1 B^{12} + \theta_1(B + \Theta_1 B^{13}))$$

$$Y_t - Y_{t-12} - \phi_1(Y_{t-1} - Y_{t-13}) - \phi_2(Y_{t-2} - Y_{t-14}) - \phi_3(Y_{t-3} - Y_{t-15}) = c + \varepsilon_t + \Theta_1 \varepsilon_{t-12} + \theta_1(\varepsilon_{t-1} + \Theta_1 \varepsilon_{t-13}))$$

$$Y_t = c + Y_{t-12} + \varepsilon_t + \Theta_1 \varepsilon_{t-12} + \phi_1(Y_{t-1} - Y_{t-13}) + \phi_2(Y_{t-2} - Y_{t-14}) + \phi_3(Y_{t-3} - Y_{t-15}) + \theta_1(\varepsilon_{t-1} + \Theta_1 \varepsilon_{t-13})$$

5.1.1.3. Estimación del modelo

De acuerdo a las diferentes pruebas realizadas, el modelo SARIMA (3,0,1) (0,1,1)₁₂ tiene buen AIC, ajuste y precisión del pronóstico que los demás modelos, es la que se estima como se muestra en la Tabla 8.

Tabla 9

Estimación de coeficientes para el modelo seleccionado

Coeficiente	Estimación
ar1	1.334
ar2	-0.485
ar3	0.117
ma1	-0.726
sma1	-0.953
Constante	0.019

Ecuación del modelo SARIMA (3,0,1) (0,1,1)₁₂ es:

$$Y_t = 0.019 + Y_{t-12} + \varepsilon_t - 0.953\varepsilon_{t-12} + 1.334(Y_{t-1} - Y_{t-13}) - 0.485(Y_{t-2} - Y_{t-14}) + 0.117(Y_{t-3} - Y_{t-15}) - 0.726(\varepsilon_{t-1} - \varepsilon_{t-13})$$

La producción de azúcar en el Perú muestra no tener una tendencia creciente significativa a largo plazo, esto porque la media de la producción es 0.019, además que la producción del mes actual depende de la producción del mismo mes del año anterior reflejando así la estacionalidad anual (Y_{t-12}), mientras que la producción actual es afectada negativamente por la producción del año anterior en el mismo mes ($-0.953\varepsilon_{t-12}$), pero también se puede observar influencias positivas de la producción por meses anteriores a la actual, por lo que se puede concluir, que el modelo predice la producción de azúcar en el Perú tomando en cuenta los patrones mensuales, la estacionalidad anual así como eventos inesperados, conllevando a planificar y gestionar la producción buscando soluciones a problemas futuros.

5.1.1.4. Evaluación del modelo

Para verificar si hay presencia de variabilidad en los residuos del modelo se utiliza el test de Breusch Pagan, donde la H_0 indica que no hay evidencia de heterocedasticidad en los residuos, y en el cual el valor de P de probabilidad de contraste es 0.1498 siendo mayor al 5% nivel de significancia, de esta forma no habiendo evidencia estadística para poder rechazar la H_0 , concluimos que no hay evidencia de heterocedasticidad en los residuos y la transformación mediante Box Cox fue exitosa en estabilizar la varianza.

De acuerdo a la prueba de Kolmogorov Smirnov para probar si los residuos siguen un comportamiento normal se encontró que, $D = 0.067828$, $p\text{-value} = 0.08136$, los datos tienen una aproximación a una distribución normal siendo el valor P de probabilidad de contraste mayor al nivel de significancia (5%) y de esta manera no mostrando evidencia suficiente para rechazar la hipótesis nula, además que la diferencia máxima entre la distribución observada y normal tiene un aproximado de 0.0678.

Mediante la prueba de Dickey Fuller, con un valor P de probabilidad de contraste 0.01, se puede constatar que, al no encontrar evidencia estadística para rechazar la hipótesis nula, se concluye que los residuos tienen un comportamiento estacionario (la media, varianza y la estructura de autocorrelación permanecen constantes en el tiempo).

Además, de acuerdo a la prueba t-student, se pudo constatar que la media de los residuos es igual a cero, esto al tener un valor P igual a 0.722, donde la hipótesis nula establece que no hay diferencias significativas entre las medias ($\mu = 0$), concluyendo con el ruido blanco sigue una distribución normal con media cero y varianza constante.

Estadística Ljung – Box, con un valor P de probabilidad de contraste 0.7899, donde la hipótesis nula (H_0 : No hay autocorrelación en los residuos, es decir, los residuos se comportan como ruido blanco).

Dado que el valor P es mayor al 5% nivel de significancia, no hay evidencias suficientes para rechazar la hipótesis nula, concluyendo que no hay autocorrelación significativa en los residuos estandarizados. Esto sugiere que el modelo que se ha ajustado (el modelo SARIMA con transformación Box Cox) es adecuado en términos de no dejar autocorrelación remanente en los residuos, lo que es un buen indicador de que el modelo ha capturado bien la estructura de los datos.

5.1.1.5. Predicción del modelo

Predicción con la data de entreno

Primero se realiza la predicción con la data de entrenamiento, tomada desde el año 1994 hasta el 2022, siendo la prueba la producción de los meses del año 2023 para su comparación.

Tabla 10

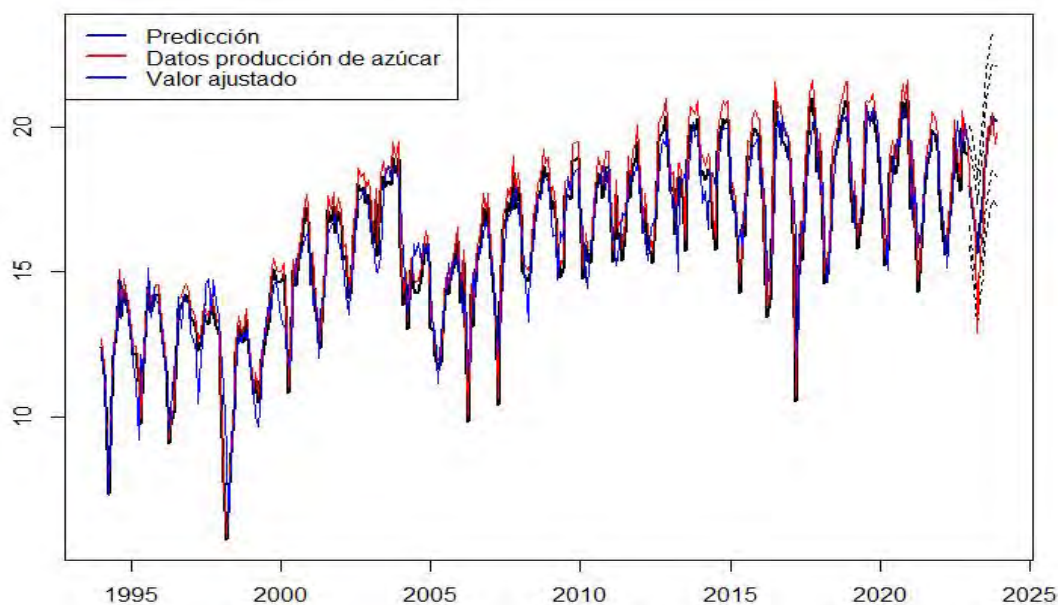
Comparación del pronóstico de la producción de azúcar y la data de prueba tomando en cuenta la transformación por Box Cox

Mes	Año	Pronóstico	Prueba transformada	Error	Li 95	LS 95
Jan	2023	17.948	17.942	0.006	19.302	20.018
Feb	2023	17.207	17.469	0.263	18.790	19.628
Mar	2023	16.423	16.483	0.060	18.066	18.936
Apr	2023	15.653	12.892	2.761	17.333	18.222
May	2023	16.665	15.690	0.976	18.380	19.287
Jun	2023	17.598	16.519	1.079	19.346	20.272
Jul	2023	18.556	18.508	0.048	20.336	21.278
Aug	2023	19.529	19.963	0.435	21.336	22.292
Sep	2023	19.702	19.541	0.162	21.534	22.504
Oct	2023	20.334	20.493	0.159	22.188	23.170
Nov	2023	20.240	19.400	0.840	22.114	23.106
Dec	2023	20.194	19.777	0.416	22.085	23.087

La Tabla 9 muestra una comparación sobre la predicción, de la producción de azúcar 1994 – 2022, con la data de entrenamiento y prueba, todo con la transformación Box Cox, tomando en cuenta la diferencia de predicción con la data de prueba transformada, así como sus límites que podría cubrir con una confianza del 95%.

Figura 17

Pronóstico y valor ajustado por el modelo SARIMA $(3,0,1)(0,1,1)_{12}$ para la serie de producción de azúcar



La Figura 17 es la representación gráfica de la predicción de la serie, así como el ajuste que tiene el modelo a los datos, en el cual se observa que capta las variaciones que presenta la serie de producción de azúcar.

Tabla 11

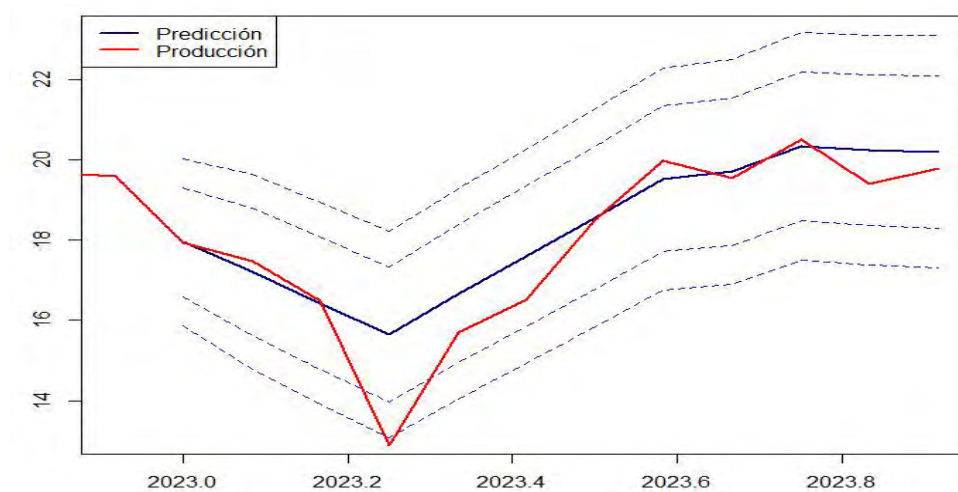
Comparación del pronóstico transformado con la data de prueba, producción de azúcar

Mes	Año	Pronóstico	Data de Prueba	Error
Jan	2023	126.165	119.429	6.736
Feb	2023	116.276	113.513	2.764
Mar	2023	106.297	101.677	4.619
Apr	2023	96.964	64.362	32.602
May	2023	109.330	92.653	16.677
Jun	2023	121.440	102.092	19.348
Jul	2023	134.599	126.727	7.873
Aug	2023	148.703	146.556	2.147
Sep	2023	151.306	140.643	10.663
Oct	2023	160.965	154.149	6.816
Nov	2023	159.508	138.702	20.805
Dec	2023	158.788	143.938	14.850

La Tabla 10 muestra los datos realizados mediante una transformación inversa para captar y comparar con los datos reales de producción, resaltando la diferencia que hay en el pronóstico con un MAE de 12.158, MAPE de 12.1% y un RMSE de 14.92.

Figura 18

Pronóstico y data de prueba de la serie producción de azúcar



La Figura 18 muestra la representación gráfica de la predicción de la serie producción de azúcar con más detalle, y se observa que la predicción trata de capturar las variaciones que la serie presenta en el tiempo, es decir que trata de seguir el patrón en el comportamiento.

5.1.2. Método de red neuronal artificial

Tomando dos grupos de la data (entrenamiento y prueba), se realiza la validación mediante una prueba con los distintos modelos de redes neuronales artificiales, tipo perceptrón multicapa, como se muestran a continuación.

Tabla 12

Comparación de modelos de redes neuronales artificiales

Modelo	Retrasos	h	Medidas de ajuste			Precisión del pronóstico		
			RMSE	MAPE	MAE	RMSE	MAPE	MAE
1	12	10	2.528	2.094	1.811	16.527	13.166	12.749
2	14	8	3.651	3.160	2.748	19.983	16.157	14.415
3	16	9	2.894	2.502	2.133	20.131	15.078	13.951
4	18	12	0.440	0.347	0.288	15.847	12.656	12.309
5	24	13	0.163	0.115	0.092	16.331	12.064	12.113

Nota. h = número de neuronas (nodos) en la capa oculta; MAPE: Error porcentual medio; RMSE: Raíz cuadrada media del error.

En la Tabla 11 se observa medidas de error, RMSE y MAPE, en los cinco modelos para el ajuste y predicción de la producción de azúcar, en el cual el modelo 5 resalta por su mejor ajuste a la serie producción de azúcar y con una capacidad de predicción menor a comparación de los demás modelos con un error porcentual de predicción (MAPE) 12.06% y de esta manera siendo elegida para realizar la predicción de la serie producción de azúcar.

Modelo de la RNA: Debido a la naturaleza de las redes neuronales y el alto grado de complejidad no lineal, su construcción del modelo de una RNA de manera explícita no es

sencilla. A diferencia de un modelo SARIMA quien se basa en una ecuación explícita y bien definida donde describe la relación entre la variable de interés y los rezagos, como los componentes estacionales y el ruido, una RNA ajusta pesos y sesgos mediante una función de activación en capas ocultas, del cual parte el problema debido a que los pesos y sesgos no muestran tener una interpretación intuitiva, porque su ajuste es mediante un proceso iterativo buscando el error mínimo.

No tiene una única ecuación, es más contiene ecuaciones unidas que dependen de las activaciones de las neuronas, las cuales contribuyen al resultado final con información que no se puede expresar en una ecuación cerrada, además puede tener miles y millones de parámetros (sesgo y peso) dependiendo de las capas y las neuronas con las que se trabaja, esto hace también una representación matemática impracticable como su interpretación en términos de la ecuación.

Tabla 13

Comparación de predicción y la data prueba de la serie producción de azúcar.

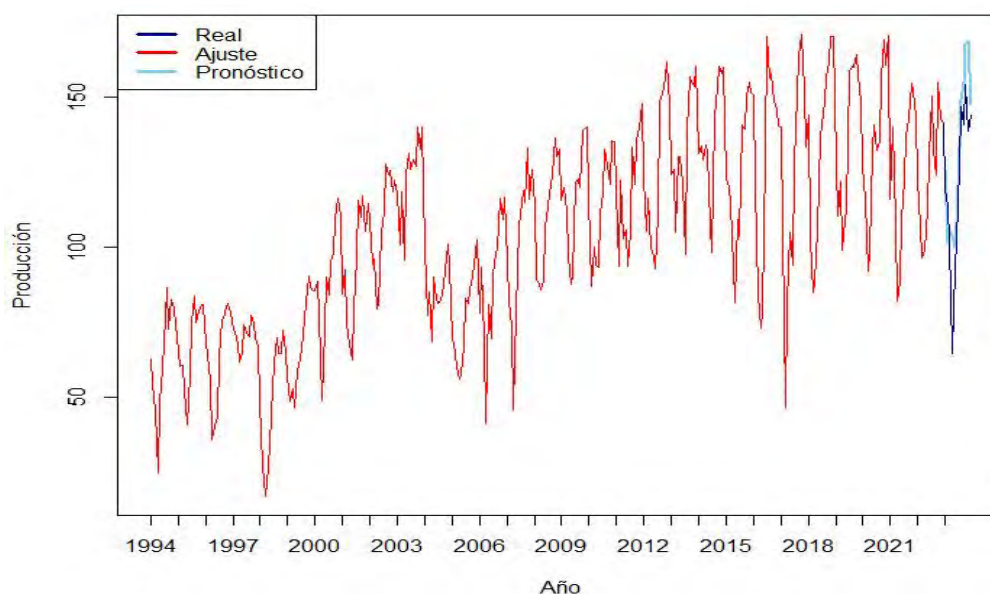
Año	Mes	Predicción	Data de Prueba	Error
2023	Jan	125.746	119.429	6.317
2023	Feb	101.249	113.513	12.264
2023	Mar	105.650	101.677	3.973
2023	Apr	103.854	64.362	39.491
2023	May	100.042	92.653	7.389
2023	Jun	108.036	102.092	5.944
2023	Jul	133.177	126.727	6.450
2023	Aug	148.858	146.556	2.301
2023	Sep	155.010	140.643	14.367
2023	Oct	167.473	154.149	13.324
2023	Nov	168.729	138.702	30.026
2023	Dec	147.449	143.938	3.510

Nota. Los valores de color negrita indican la diferencia de error mayor a 10.

En la Tabla 12 se observa una comparación de los valores predichos con la data de entrenamiento y prueba, mostrando que en los meses febrero, abril, setiembre, octubre y noviembre la diferencia entre la predicción y el valor real de la producción son altos, llegando a concluir que el modelo no aprendió lo suficientemente bien para considerar dichos meses en su ajuste.

Figura 19

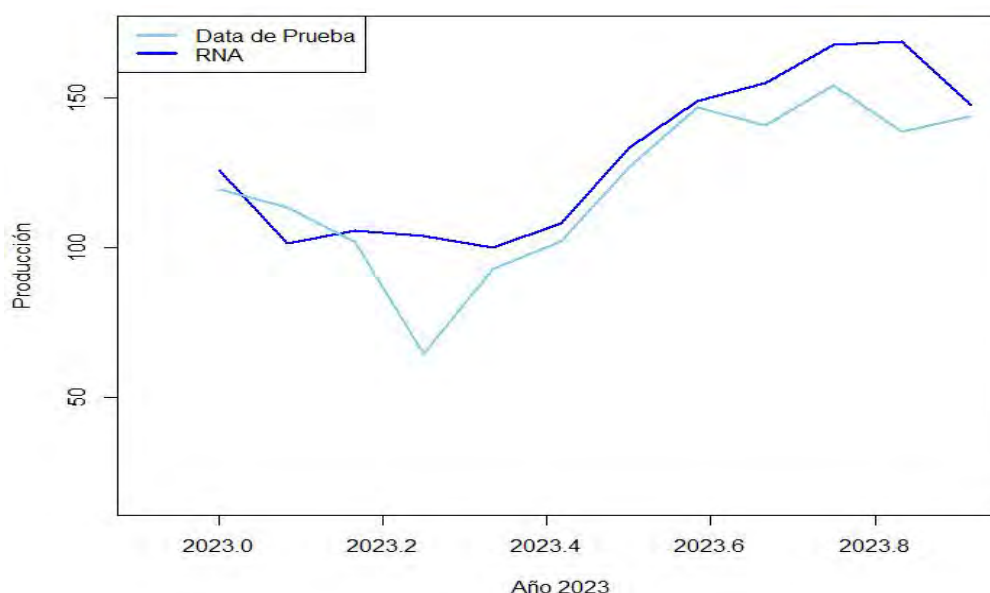
Predicción y valor ajustado por la RNA para la producción de azúcar



En la Figura 19 se observa la predicción y su ajuste a los datos reales, mostrando que el modelo se ajusta a los datos de mejor manera, pero la predicción está un poco más alejada de la data.

Figura 20

Comparación de la predicción con la data de prueba, producción de azúcar



La Figura 20 muestra una comparación gráfica de la predicción con la data de prueba, en el cual se puede observar que la predicción trata de seguir el patrón que tiene el comportamiento de la serie producción de azúcar, sin embargo, muestra cometer errores visualmente grandes.

5.1.3. Comparación de método Box – Jenkins y RNA

Tabla 14

Medidas de ajuste del método Box – Jenkins y la red neuronal artificial

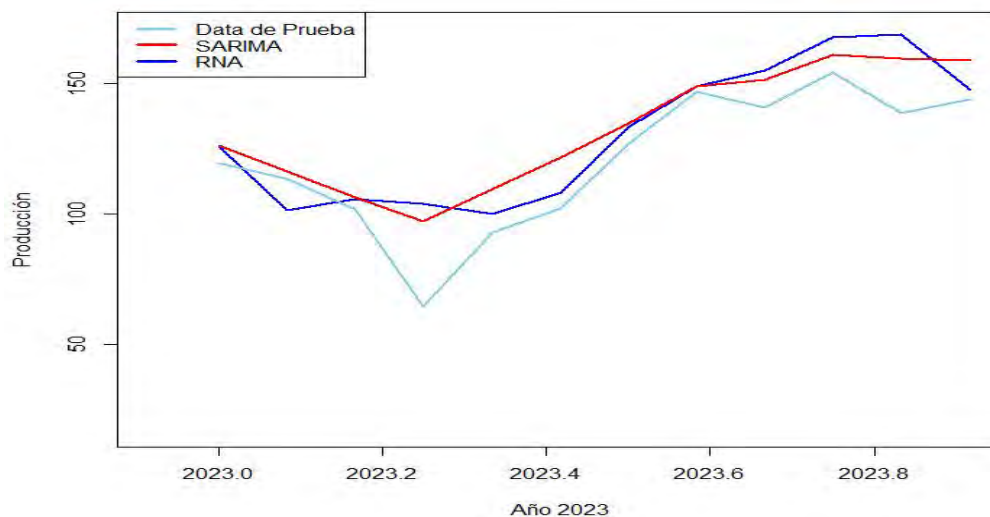
Método	Modelo	Medidas de ajuste			Precisión del pronóstico		
		RMSE	MAPE	MAE	RMSE	MAPE	MAE
Box - Jenkins	SARIMA(3,0,1)(0,1,1) ₁₂	1.025	5.095	0.760	0.955	3.879	0.600
RNA	5 24 (13)	0.163	0.115	0.092	16.331	12.064	12.113

La Tabla 13 muestra los modelos escogidos en ambas metodologías con los datos de entrenamiento, y de acuerdo al RMSE y MAPE se puede concluir que el modelo 5 de RNA

se ajusta mejor a los datos, pero el modelo SARIMA es quien tiene mejor precisión con respecto a la predicción.

Figura 21

Gráfico de comparación SARIMA, RNA y los datos reales del año 2023



En la Figura 21 se observa una comparación visual de la predicción entre ambos modelos con el valor real de la producción de azúcar para el año 2023, donde el modelo SARIMA $(3, 0, 1)(0, 1, 1)_{12}$ trata de capturar y seguir el patrón que sigue el comportamiento de la serie, mientras que el modelo de RNA en ciertos momentos se aleja como también se acerca al comportamiento de la serie producción de azúcar.

DISCUSIÓN

En los estudios realizados por Miranda Chinlli (2021), cuyo objetivo general es presentar la metodología de Box Jenkins para modelos SARIMA, que ayude a pronosticar la temperatura de la ciudad de New York, utilizando datos históricos de series de tiempo, recopilados de la estación meteorológica por el proyecto CRUTEM, concluye que la metodología de Box y Jenkins otorga una manera sencilla para la predicción de series de tiempo, por su eficacia de ajuste y su rapidez, además que es útil a corto plazo; de la misma manera Ruiz, Hernández y Zulueta (2011), en su investigación “Análisis de series de tiempo en el pronóstico de la producción de caña de azúcar”, donde el objetivo es preveer el comportamiento de su proceso productivo con el fin de planificar y optimizar el uso de los recursos técnicos, humanos y financieros, concluyen que el modelo de Box Jenkins, ARIMA (1, 2, 0), es muy bueno en la predicción de series respecto a producción, logrando una precisión del 94%.

En la investigación de Menacho (2013), cuyo objetivo es comparar los métodos estadísticos clásicos aplicados al análisis de series de tiempo con los modelos de redes neuronales artificiales multicapa back propagación, concluye con que los modelos de redes neuronales multicapa muestran resultados favorables por la precisión y capacidad de aprender además de que se adaptan rápido conllevando una ventaja respecto a modelos clásicos de series de tiempo; mientras que Lazo Chuquihuayta (2019), cuyo objetivo general es comparar el modelo de redes neuronales artificiales para el pronóstico de número de visitantes extranjeros a Machu Picchu 2002 – 2018 con la metodología Box Jenkins, llega a la conclusión que la RNA frente al modelo de Box y Jenkins (ARIMA), captura de mejor forma el comportamiento de los datos y por ende su mejor precisión del pronóstico; el

estudio realizado por Puma Cárdenas (2020), donde el objetivo principal es analizar los modelos autorregresivos integrado de media móvil (ARIMA) o la red neuronal recurrente tipo de memoria a corto y largo plazo (RNNLSTM) para predecir adecuadamente la demanda de medicamentos en el hospital regional del Cusco, 2018, llego a la conclusión de que la RNNLSTM predice de mejor manera frente a un ARIMA, esto basándose en los resultados del MAPE con un bajo porcentaje de error cometido.

Tomando en cuenta los resultados del presente estudio y los antecedentes ya mencionados, se encuentra cierta similitud en los resultados, alegando que la metodología de Box y Jenkins es útil a series de corto plazo y las redes neuronales son eficaces al adaptarse y aprender rápido, y en dicha investigación se pudo observar que el modelo de las redes neuronales tiene la capacidad de aprender y ajustarse a los datos, mientras que el modelo SARIMA de la metodología Box Jenkins tuvo la capacidad de ajustar la predicción con mayor precisión que el modelo de red neuronal artificial, con un MAPE en promedio 3.9% y 12.1% respectivamente en la precisión del pronóstico, como se muestra en la Tabla 13.

CONCLUSIONES

El modelo de la metodología Box-Jenkins proporciona una precisión en el pronóstico de la serie producción de azúcar, SARIMA (3, 0, 1) (0, 1, 1)₁₂, con un RMSE de 0.95, MAE de 0.6 y un MAPE de 3.9%.

El modelo de la red neuronal artificial de tipo perceptrón multicapa con 13 neuronas en la capa oculta, muestra un mejor ajuste a la serie, mientras que la precisión de pronóstico no es tan buena como su ajuste, teniendo un RMSE de 16.3, MAE igual a 12.1 y un MAPE de 12.1%.

En la serie de la producción de azúcar, el modelo 5 de la red neuronal artificial muestra un mejor ajuste a la serie producción de azúcar, sin embargo, el modelo SARIMA (3, 0, 1) (0, 1, 1)₁₂ de la metodología Box Jenkins es quien resalta en tener mejor precisión al momento de realizar la predicción, con un error menor cometido.

Concluyendo, que el resultado contradice la hipótesis general, que proponía un mejor desempeño de la red neuronal artificial resaltando la metodología Box Jenkins, siendo el modelo escogido para la predicción de la serie de producción del azúcar en el Perú.

RECOMENDACIONES

Realizar predicciones utilizando diferentes series sobre producción de alimentos de primera necesidad o productos, y también sobre series con presencia de alta variabilidad, de esta manera comparar las redes neuronales y los modelos clásicos de series de tiempo, encontrando ventajas y desventajas a mejorar.

Buscar algoritmos óptimos que entrenen los datos mientras se ejecutan, o algoritmos que no demoren en el proceso de entrenamiento.

En cuanto a las redes neuronales, son más eficientes mientras más datos de entrenamiento tengan, por ello utilizar series con mayor tamaño de observaciones y variables.

REFERENCIAS

- Abarca Calderón, F. (s.f.). Modelos probabilísticos de señales y sistemas. Kerwa. Kerwa:
<https://www.kerwa.ucr.ac.cr/server/api/core/bitstreams/e213bb92-3358-4b46-8b96-d3cbf0c2c517/content>
- Agroindustrial Laredo S. A. A. (Febrero de 2024). *Cutivamos cosas buenas*. Memoria Anual 2023, Sol de Laredo. <https://agroindustriallaredo.com/wp-content/uploads/2025/03/Memoria-del-Directorio-2024.pdf>
- Asociación Peruana de Productores de Azúcar. (2002).
<https://www2.congreso.gob.pe/sicr/tradocestproc/clproley2001.nsf/pley/DCE228F5012FFD6D05256D25005D33F8?opendocument>
- Ballat, M., Chavarrias Navarro, Á., & Purroy Arnó, E. (2014). Valorización de subproductos en la industria azucarera y análisis del ciclo de vida. *Universidad Pública de Navarra*.
https://www.researchgate.net/publication/304989571_Valorizacion_de_subproductos_en_la_industria_azucarera_y_Analisis_del_Ciclo_de_Vida/figures?lo=1
- Cáceres Hernández, J. J. (2006). Variable aleatoria real. En *Conceptos básicos de estadística para ciencias sociales* (págs. 171-173).
<https://bibliotecaenlinea.universidaddorados.edu.mx/contaduriapublica/Estadistica/Conceptos%20estadistica.pdf>
- Comité Nacional para el Desarrollo Sustentable de la caña de Azúcar. (1 de Enero de 2016). *Información que es importante de saber sobre el consumo de azúcar*. Gobierno de México:
<https://www.gob.mx/conadesuca/documentos/informacion-que-es-importante-saber-sobre>

el-consumo-de-azucar-

233909?state=published#:~:text=El%20az%C3%BAcar%20ha%20sido%20considerado,cerebro%20como%20al%20sistema%20nervioso.

CONADESUCA. (Enero de 2015). *Comité Nacional para el Desarrollo Sustentable de la Caña de Azúcar.*

https://www.gob.mx/cms/uploads/attachment/file/141823/Ficha_Tecnica_Ca_a_de_Azucar.pdf

De la Guía Gonzales, A. V. (2017). *La transformación de Box - Cox*. Universidad Complutense Madrid, Facultad de Ciencias Económicas y Empresariales, Madrid.

Dip, J. A., & Romero, P. I. (2015). Una comparación de redes neuronales y modelos arch-garch para predecir variaciones en el precio de acciones. aplicación a un caso de acciones de telefonía. 2. <https://doi.org/2250-6861>

Enciclopedia Británica. (15 de Noviembre de 2024). *Britannica*. (V. Srinivasa, Editor) Britannica Web site: <https://www.britannica.com/science/random-walk>

Enciclopedia Humanidades. (23 de Enero de 2023). (Equipo editorial, Etecé) Cerebro: partes, funciones, características y enfermedades: <https://humanidades.com/cerebro/>

Estadísticas Cómo hacerlo. (s.f.). Statistics How To Web site: <https://www.statisticshowto.com/probability-and-statistics/normal-distributions/box-cox-transformation/>

- González Casimiro, M. P. (2009). *Análisis de series temporales: Modelos ARIMA*. (F. d. Empresariales, Ed.) Universidad del País Vasco. <https://doi.org/978-84-692-3814-1>
- Gutierrez Ramos, D. (Junio de 2022). Análisis de series temporales: Estudio de un caso práctico. https://oa.upm.es/70926/1/TFG_DAVID_GUTIERREZ_RAMOS.pdf
- Hilera Gonzales, R. J., & Martínez Hernando, V. J. (1995). *Redes Neuronales Artificiales: fundamentos, modelos y aplicaciones*. Madrid, España: RA-MA Editorial. https://www.researchgate.net/profile/Jose_Hilera/publication/44343683_Redes_neuronales_artificiales_fundamentos_modelos_y_aplicaciones_Jose_Ramon_Hilera_Gonzalez_Victor_Jose_Martinez_Hernando/links/5886033892851c21ff4d3060/Redes-neuronales-artificiales-fu
- InfoAgro. (2022). Descripción del cultivo de caña de azúcar. *Revista InfoAgro México*. <https://mexico.infoagro.com/descripcion-del-cultivo-de-cana/>
- Isasi Viñuela, P., & Galvan León, I. M. (2004). *Redes de neuronas artificiales un enfoque práctico*. (D. F. Aragón, Ed.) Madrid, España: Pearson Educación, S.A. <https://doi.org/ISBN: 84-205-4025-0>
- Laurentin, H. (2020). *Importancia de la predicción del rendimiento en caña de azúcar en un contexto de transformación digital*. SofOS Corporation web site: <https://sofoscorp.com/2020/02/importancia-de-la-prediccion-del-rendimiento-en-cana-de-azucar-en-transformacion-digital/>
- Lazo Chuquihuayta, H. (2019). *Modelo de redes neuronales artificiales para el pronóstico del número de visitantes extranjeros a Machu Picchu en comparación con la metodología de*

box y jenkins. Tesis en maestria, Universidad Nacional de San Antonio Abad del Cusco, Cusco, Perú. <http://purl.org/pe-repo/ocde/ford#1.01.03>

Maté Jimenez, C. (1995). *Variables aleatorias y distribuciones de probabilidad*. https://www.google.com.pe/books/edition/Curso_general_sobre_Statgraphics/7X-csZNqpm0C?hl=es&gbpv=1&dq=variable+aleatoria+discreta&pg=PA395&printsec=frontcover

Mauricio, J. A. (2007). *Análisis de series temporales*. Madrid, España: Universidad Complutense de Madrid. https://gc.scalahed.com/recursos/files/r161r/w25599w/L1EF118_MAURICIO.pdf

Menacho Ch., C. (2013). Comparación de los métodos de series de tiempo y redes neuronales. *Revista Universidad Nacional Agraria La Molina*. <https://doi.org/ISSN:2519-7398>

MIDAGRI. (2023). *Azúcar*. Ministerio de Desarrollo Agrario y Riego, Dirección de estudios económicos. Lima: Dirección General de Políticas Agrarias. https://repositorio.midagri.gob.pe/bitstream/20.500.13036/1581/1/Commodities%20Az%C3%BAcar_%20abr-jun%202023%20%281%29.pdf

Ministerio de Agricultura, Pesca y Alimentación. (s.f). *Azúcar*. https://www.mapa.gob.es/es/ministerio/servicios/informacion/azucar_tcm30-102346.pdf

Miranda C., R., Marza M., R., Calle P., J. L., Choque T., C., Mendoza A., M., Cruz Ch., D., & Aparicio P., J. J. (2021). Aptitud del uso del suelo para el cultivo de la caña de azúcar (*Saccharum officinarum* L.) en la localidad de Tumupasa, Municipio de San Buenaventura

- La paz. *Revista de Investigación e Innovación Agropecuaria y de Recursos Naturales*, vol. 8(1), 25 - 36. <https://doi.org/ISSN: 2518-6868>

Miranda Chinlli, C. (2021). *Modelización de Series Temporales modelos clásicos y SARIMA*. Tesis de Maestría, Universidad de Granada, Departamento de Estadística e investigación Operativa, Granada, España. https://masteres.ugr.es/estadistica-aplicada/sites/master/moea/public/inline-files/TFM_MIRANDA_CHINLLI_CARLOS.pdf

PERUCANA. (2023). *Informe anual*. Asociación Peruana de Agroindustriales: del Azúcar y Derivados. Industria Azucarero. https://www.perucana.com/_files/ugd/43b6dc_2a9b65412c864234a33ca16e1ff5c6f0.pdf

Puma Cárdenas, A. (2020). *Modelos Autorregresivos Integrados de Media Móvil y Red Neuronal Recurrente para predecir la demanda de medicamentos, hospital regional Cusco, 2013 - 2018*. Cusco, Perú.

Ríos, G. (2008). *Series de Tiempo*. Universidad de Chile, Departamento de Ciencias de la Computación. Facultad de Ciencias Físicas y Matemáticas. https://www.u-cursos.cl/ingenieria/2010/1/CC52A/1/material_docente/bajar?id_material=296003

Ruiz Ramírez, J., Hernández Rodríguez, G. E., & Zulueta Rodríguez, R. (2011). Análisis de series de tiempo en el pronóstico de la producción de caña de azúcar. 29(1). https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-57792011000100103

Russell, R. (2018). *Redes Neuronales: Guía sencilla de las redes neuronales artificiales*. https://books.google.com/books/about/Redes_Neuronales.html?id=3OtcuQEACAAJ

Sol de Laredo S. A. A. (Febrero de 2024). *Cutivamos cosas buenas*.

Villavicencio, J. (2010). *Introducción a series de tiempo*.

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

ANEXOS

a. Matriz de consistencia

Problemas General	Objetivo General	Hipótesis General	Variables e indicadores	Metodología
¿Cuál de los modelos, metodología Box Jenkins y las redes neuronales artificiales permite una mejor predicción de la producción de azúcar?	Analizar los modelos de la metodología Box Jenkins y la red neuronal artificial para encontrar la predicción adecuada de la producción de azúcar en el Perú.	El modelo de la red neuronal artificial permite una mejor predicción de la producción de azúcar en el Perú.	<p>Variable de estudio dependiente</p> <p>Producción de azúcar.</p> <p>Variable de estudio independiente</p> <p>Tiempo (1994 – 2023)</p>	<p>Enfoque:</p> <p>Cuantitativo</p> <p>Diseño:</p> <p>Es una investigación no experimental y transversal de acuerdo a la finalidad de la misma.</p> <p>Población:</p> <p>La población de estudio comprende la secuencia de la cantidad de producción total mensual del azúcar, recolectada desde el mes de enero de 1994 hasta el</p>

				<p>mes de diciembre de 2023, los datos se encuentran en la base de datos del Banco Central de Reserva del Perú (BCRP), donde están ordenados y equidistantes en diferentes momentos en el tiempo.</p>
--	--	--	--	---

<p>Específicos</p> <p>¿Cuál es el pronóstico para la producción de azúcar utilizando modelos de redes neuronales artificiales?</p> <p>¿Cuál es la predicción la producción de azúcar utilizando modelos de series de tiempo utilizando la metodología de Box - Jenkins?</p> <p>¿Cuál de los modelos realiza una mejor predicción para la producción de azúcar</p>	<p>Específicos</p> <p>Determinar el pronóstico de la producción de azúcar utilizando modelos de redes neuronales artificiales.</p> <p>Determinar la predicción de la producción de azúcar utilizando modelos de series de tiempo utilizando la metodología de Box - Jenkins.</p> <p>Comparar y evaluar la mejor predicción de los modelos de redes neuronales artificiales</p>	<p>Específicos</p> <p>Los modelos SARIMA, de la metodología Box y Jenkins, proporcionan un mejor ajuste para la predicción de la producción de azúcar en el Perú.</p> <p>Los modelos de red neuronal artificial, tipo perceptrón multicapa, proporcionan un mejor ajuste para la predicción de la producción de</p>	<p>Técnicas e instrumentos para la recolección de datos:</p> <p>En dicho trabajo las principales técnicas que se empleó fue la recolección de datos mediante una búsqueda en internet en la base de datos del BCRP.</p> <p>Métodos:</p> <p>Redes neuronales artificiales y series de tiempo.</p>
--	---	--	--

utilizando redes neuronales artificiales y los modelos de series de tiempo?	con los modelos de series de tiempo.	<p>azúcar en el Perú.</p> <p>El modelo SARIMA y la red neuronal artificial presentan diferencias en su capacidad de ajuste y predicción, siendo uno de ellos adecuado para realizar predicciones más precisas.</p>		
---	--------------------------------------	--	--	--

b. Otros

Serie mensual de producción de azúcar en el Perú

Año	Meses											
	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
1994	62.512	54.555	45.219	24.724	47.943	59.782	67.659	86.300	72.470	82.512	79.311	74.404
1995	65.205	60.238	60.430	52.342	40.516	57.202	74.744	83.554	74.531	78.435	80.646	80.597
1996	67.918	63.926	55.402	35.753	40.109	43.185	63.042	75.717	76.198	79.879	81.062	77.835
1997	72.562	71.361	67.606	61.483	64.287	74.029	71.492	69.985	77.285	74.613	69.657	67.027
1998	48.499	27.112	16.808	21.138	33.396	44.289	61.046	69.695	64.417	64.327	72.272	64.879
1999	56.457	48.420	52.648	46.310	58.808	60.061	66.083	70.728	81.346	90.323	86.377	85.290
2000	85.398	88.465	77.041	48.744	72.479	89.861	83.900	95.262	97.788	108.963	116.159	110.217
2001	84.314	92.509	74.838	67.561	62.152	79.394	96.293	115.337	109.813	116.844	105.116	114.452
2002	108.693	94.099	95.921	79.334	80.880	104.381	111.578	127.367	124.003	125.455	118.260	122.081
2003	118.035	100.705	118.258	95.548	124.179	130.998	125.604	128.778	126.890	139.934	132.381	139.803
2004	99.367	77.117	85.091	68.406	90.040	82.982	81.213	81.970	87.019	92.066	100.791	93.825
2005	69.110	68.068	60.083	55.902	57.660	66.783	82.989	81.077	85.753	90.364	98.019	102.393
2006	77.829	93.247	69.506	40.978	80.829	69.468	90.714	97.764	99.465	116.230	108.953	116.445
2007	94.941	81.848	73.730	45.673	78.176	106.903	110.842	118.760	114.705	132.924	115.877	125.621
2008	113.724	89.380	87.440	85.655	88.546	106.140	113.534	120.879	123.911	136.272	129.994	132.352
2009	115.648	119.636	111.785	96.581	87.375	90.427	121.632	122.406	119.893	138.586	139.236	140.017
2010	116.173	86.621	100.028	93.838	92.968	111.361	117.788	132.586	127.323	120.769	135.383	135.140
2011	119.055	93.332	122.207	102.542	105.420	93.654	105.171	133.254	120.909	135.673	139.494	147.879
2012	123.028	105.110	116.328	99.911	96.623	92.680	110.731	148.048	151.226	152.857	161.546	147.247
2013	123.973	125.838	104.748	130.017	129.999	121.037	97.498	138.983	156.701	155.088	153.493	160.203

2014	131.193	133.428	128.809	134.036	131.456	107.051	98.035	143.085	153.143	160.019	157.597	159.763
2015	122.719	120.096	114.210	93.141	81.430	109.000	103.548	140.649	139.158	150.297	154.994	151.201
2016	150.698	120.575	84.288	72.747	78.584	130.556	170.082	156.403	157.657	148.487	146.064	140.241
2017	140.108	105.290	46.489	80.567	104.969	93.770	127.890	151.111	163.416	170.942	152.444	133.344
2018	143.802	102.601	84.762	88.224	105.784	136.884	140.787	148.568	156.714	161.641	169.945	170.021
2019	127.864	109.987	121.786	98.718	108.728	112.775	158.539	160.103	159.630	163.945	157.229	148.121
2020	127.112	115.398	91.872	97.931	132.576	140.657	132.199	135.297	156.300	169.002	160.460	170.481
2021	115.675	139.821	107.870	81.692	88.673	109.261	122.417	136.831	142.852	154.401	151.872	143.520
2022	122.278	106.841	96.312	98.757	112.695	130.021	150.079	132.530	123.932	154.862	142.830	141.372
2023	119.429	113.513	101.677	64.362	92.653	102.092	126.727	146.556	140.643	154.149	138.702	143.938

Fuente: Banco Central de Reserva del Perú (BCRP).