Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
Facultad de Ingeniería Industrial
Escuela Profesional de Ingeniería Industrial

Influencia de la aplicación de técnicas econométricas de series de tiempo y Machine Learning en el pronóstico de la demanda de cursos de la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM

TESIS

Para optar el Título Profesional de Ingeniero Industrial

AUTOR

Daniel CORILLOCLLA HUAMAN

ASESOR

Alfonso Ramón CHUNG PINZÁS

Lima, Perú

2025



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

https://creativecommons.org/licenses/by-nc-sa/4.0/

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Corilloclla, D. (2025). *Influencia de la aplicación de técnicas econométricas de series de tiempo y Machine Learning en el pronóstico de la demanda de cursos de la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería Industrial, Escuela Profesional de Ingeniería Industrial]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Daniel Corilloclla Huaman
Tipo de documento de identidad	DNI
Número de documento de identidad	47491081
URL de ORCID	https://orcid.org/0009-0001-2002-825X
Datos de asesor	
Nombres y apellidos	Alfonso Ramón Chung Pinzás
Tipo de documento de identidad	DNI
Número de documento de identidad	09394397
URL de ORCID	https://orcid.org/0000-0002-8073-2396
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Juan Manuel Cevallos Ampuero
Tipo de documento de identidad	DNI
Número de documento de identidad	07855059
Miembro del jurado 1	
Nombres y apellidos	Oscar Rafael Tinoco Gómez
Tipo de documento de identidad	DNI
Número de documento de identidad	08606920
Miembro del jurado 2	
Nombres y apellidos	Daniel Víctor Surco Salinas
Tipo de documento de identidad	DNI
Número de documento de identidad	09722150
Datos de investigación	
Línea de investigación	1 Inteligencia Artificial. ODS 4: EDUCACIÓN DE CALIDAD. C.17.0.1
Grupo de investigación	No Aplica
Agencia de financiamiento	Sin financiamiento

Ubicación geográfica de la investigación	Edificio: Universidad Nacional mayor de San Marcos País: Perú Departamento: Lima Provincia: Lima Distrito: Cercado de Lima Calle: Germán Amezaga 325 Latitud: -12.05819 Longitud: -77.01891
Año o rango de años en que se realizó la investigación	Enero 2024 – marzo 2025
URL de disciplinas OCDE	Ingeniería industrial https://purl.org/pe-repo/ocde/ford#2.11.04 Educación general https://purl.org/pe-repo/ocde/ford#5.03.01 Estadísticas, Probabilidad https://purl.org/pe-repo/ocde/ford#1.01.03

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, Decana de América

FACULTAD DE INGENIERÍA INDUSTRIAL

VICEDECANATO ACADÉMICO

ACTA DE SUSTENTACIÓN Nº033-VDAP-FII-2025

SUSTENTACIÓN DE TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO INDUSTRIAL

El Jurado designado por la Facultad de Ingeniería Industrial, reunido en acto público en el Auditorio de la Facultad de Ingeniería Industrial, el día **lunes 09 de junio de 2025**, a las 12:00 horas, dio inicio a la sustentación de la tesis:

"INFLUENCIA DE LA APLICACIÓN DE TÉCNICAS ECONOMÉTRICAS DE SERIES DE TIEMPO Y MACHINE LEARNING EN EL PRONÓSTICO DE LA DEMANDA DE CURSOS DE LA ESCUELA PROFESIONAL DE INGENIERÍA TEXTIL Y CONFECCIONES DE LA UNMSM"

Que presenta el Bachiller:

DANIEL CORILLOCLLA HUAMAN

Para optar el Título Profesional de Ingeniero Industrial en la Modalidad: **Ordinaria.**

Luego de la exposición, absueltas las preguntas del Jurado y siendo las 13h00 horas se procedió a la evaluación secreta, habiendo sido APROBADO por UNANIMIDAD con la calificación promedio de DIECISIETE, lo cual se comunicó públicamente.

Lima, 09 de junio del 2025

DR. JUAN MANUEL CEVALLOS AMPUERO
Presidente

DR. OSCAR RAFAEL TINOCO GÓMEZ
Miembro

DR. DANIEL VÍCTOR SURCO SALINAS
Miembro

DR. ALFONSO RAMÓN CHUNG PINZÁS
Asesor

MG. LUIS ROLANDO RAEZ GUEVARA
Vicedecano Académico – FII

UNIVERSIDAD NACIONAL MAYOR DE
SAN MARCOS
Universidad del Perú, Decana de América

CERTIFICADO DE SIMILITUD

Yo, **CHUNG PINZÁS ALFONSO RAMÓN** en mi condición de asesor acreditado con **DICTAMEN N° 001-EPII-FII-2025** de la tesis, cuyo título es **INFLUENCIA DE LA APLICACIÓN DE TÉCNICAS ECONOMÉTRICAS DE SERIES DE TIEMPO Y MACHINE LEARNING EN EL PRONÓSTICO DE LA DEMANDA DE CURSOS DE LA ESCUELA PROFESIONAL DE INGENIERÍA TEXTIL Y CONFECCIONES DE LA UNMSM**, presentado por el bachiller **CORILLOCLLA HUAMAN DANIEL** para optar el **TÍTULO PROFESIONAL DE INGENIERO INDUSTRIAL**, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el **11%** de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional.**

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del Título Profesional especialidad correspondiente.

Lima, 07 de agosto del 2025



Firma del Asesor
Nombres y Apellidos del asesor: **CHUNG PINZÁS ALFONSO RAMÓN**
DNI: **09394397**

RESUMEN

La presente investigación presenta el análisis de la serie de datos en las matrículas en los semestres comprendidos en periodos 2010–0 al 2021-I dentro de la Universidad Nacional Mayor de San Marcos de la Escuela Profesional de Ingeniería Textil y Confecciones donde se evalúa la calidad de pronóstico con el fin de mejorar el pronóstico en la cantidad de vacantes requeridos por los estudiantes al momento de iniciar nuevos ciclos académicos. El propósito de analizar la calidad de pronóstico beneficiará a resolver diversos cuellos de botella que se generan en el proceso operativo de la matricula, en la gestión por cursos, asignación de docentes, entre otros que se generan por la baja calidad de pronóstico. Se evaluaron tres métodos econométricos (ARIMA, Holt-Winters, Redes Neuronales) en el que se aplicaron tres estadísticos de prueba a los residuales de la serie: son cercanos a cero es decir siguen una distribución normal, están dentro de los intervalos de confianza de ACF y están distribuidos independientemente con la cual se comprueba si son métodos viables para la serie para así comprobar las hipótesis. En resultaos se comprueba de los tres modelos, el modelo Holt-Winters no es aplicable a la serie, sin embargo, los métodos ARIMA y las Redes Neuronales si superan las pruebas estadísticas, así pues, se concluye que el modelo ARIMA es el que tiene mayor calidad en pronostico debido a que el estadístico de prueba de error RMSE es menor siendo la mejor opción como método para la presente investigación.

Palabras claves: Series de tiempo, Redes Neuronales, Pronóstico, Machine Learnig.

ABSTRACT

The present investigation presents the analysis of the data series in the enrollments in the semesters included in the periods 2010-0 to 2021-I within the Universidad Nacional Mayor de San Marcos of the Professional School of Textile Engineering and Confections where the quality is evaluated. in order to improve the forecast in the number of vacancies required by students at the time of starting new academic cycles. The purpose of analyzing the forecast quality will benefit to solve various bottlenecks that are generated in the operative process of enrollment, course management, teacher assignment, among others that are generated by the low forecast quality. Three econometric methods (ARIMA, Holt-Winters, Neural Networks) were evaluated in which three test statistics were applied to the residuals of the series: they are close to zero, that is, they follow a normal distribution, they are within the confidence intervals of ACF and are distributed independently with which it is verified if they are viable methods for the series in order to verify the hypotheses. In results, it is verified of the three models, the Holt-Winters model is not applicable to the series, however, the ARIMA methods and the Neural Networks do pass the statistical tests, so it is concluded that the ARIMA model is the one with the highest quality. in prognosis because the RMSE error test statistic is lower, being the best option as a method for the present investigation.

Keywords: Time series, Neural Networks, Forecast, Machine Learning.

ACTA DE SUSTENTACIÓN

DEDICATORIA

A mis padres Julián Corilloclla y Dina Huaman, a mi hermano Luis, a mis primos, a mis tíos, a mis amigos de la universidad.

AGRADECIMIENTOS

A nuestro señor por todas las bendiciones.

A mi asesor de tesis, Dr. Alfonso Chung Pinzás, quien con su valioso apoyo permitió mejorar la presente tesis.

A mis docentes, por el gran aporte a mi persona como profesional.

INDICE GENERAL

RESUMEN i

ABSTRACT ii

INDICE DE CUADROS ix

INDICE DE FIGURAS xi

CAPÍTULO I: INTRODUCCIÓN 1

1.1. Introducción 1

1.2. Planteamiento del Problema 2

1.2.1. Determinación del problema 2

1.2.2. Formulación del problema 4

1.3. Objetivos 5

1.3.1. Objetivo principal 5

1.3.2. Objetivos secundarios 5

1.4. Importancia y alcance de la investigación 6

1.5. Limitaciones de la investigación 6

CAPÍTULO II: REVISIÓN DE LITERATURA 7

2.1. Marco teórico 7

2.2. Antecedentes del estudio 7

2.2.1. Internacional 7

2.2.1.1. Nacional 8

2.3. Bases teóricas 10
2.3.1. Series de tiempo 10
2.3.2. Dominios de una serie de tiempo 14
2.3.3. Componentes de una serie de tiempo 14
2.3.4. Redes Neuronales 19
2.3.5. ARIMA 26
2.3.6. Holt-Winters 30
2.3.7. Lenguaje de programación R 33
2.3.8. Rstudio 35
2.4. Definición de términos 43
CAPÍTULO III: HIPÓTESIS Y VARIABLES 45
3.1. Hipótesis 45
3.1.1. Hipótesis principal 45
3.1.2. Hipótesis secundarias 45
3.2. Variables 46
3.4. Operacionalización de las variables 47
3.5. Matriz de consistencia 48
CAPÍTULO IV: MATERIALES Y MÉTODOS 50
4.1. Área de estudio 50
4.2. Tipo y diseño de investigación 50
4.3. Población, muestra y unidad de análisis 50

4.4. Procedimientos, técnicas e instrumentos de recolección de datos.....51
4.5. Análisis estadístico 51
CAPÍTULO V: RESULTADOS 52
5.1. Análisis de los patrones de la serie de tiempo 52
5.1.1. Curso de algoritmos y programación.....52
5.1.2. Curso de Calculo III 54
5.2. Aplicación de modelos.....56
5.2.1. Curso de Algoritmos y programación 57
5.2.2. Curso de Calculo III 79
5.3. Prueba de hipótesis.....101
5.3.1. Hipótesis secundaria 1.....101
5.3.2. Hipótesis secundaria 2.....101
5.3.3. Hipótesis secundaria 3.....102
5.3.4. Hipótesis general 103
CAPÍTULO VI: DISCUSIÓN 104
CAPÍTULO VII: CONCLUSIONES Y RECOMENDACIONES 109
7.1. Conclusiones 109
7.2. Recomendaciones.....111
REFERENCIAS BIBLIOGRÁFICAS 114
ANEXOS 118

ANEXO 1 – Matrículas del curso de Algoritmos y Programación en los periodos 2010-I al 2021-I. 118

ANEXO 2 – Matrículas del curso de Calculo III en los periodos 2010-I al 2021-I. 120

ANEXO 3 – SCRIPT aplicado al curso de algoritmo y programación 123

ANEXO 4 – SCRIPT aplicados al curso de cálculo III 128

ANEXO 5 – PRONOSTICOS del curso de cálculo III 133

ANEXO 6 – PRONOSTICOS del curso Algoritmos y programación 135

INDICE DE CUADROS

Cuadro 1 Operacionalización de las variables 47

Cuadro 2 Matriz de consistencia 48

Cuadro 3 Tabla de pronóstico por el modelo de redes neuronales para cinco periodos del curso Algoritmos y programación según ANEXO 6. 61

Cuadro 4 Tabla de pronóstico por el modelo redes neuronales para veintidós periodos del curso algoritmos y programación según ANEXO 6. 62

Cuadro 5 Tabla de pronóstico por el modelo ARIMA para cinco periodos del curso Algoritmos y programación, según ANEXO 6 66

Cuadro 6 Tabla de pronóstico por el modelo ARIMA para veintidós periodos del curso Algoritmos y programación según ANEXO 6. 68

Cuadro 7 Tabla de pronóstico para cinco periodos del curso de algoritmos y programación por el método HOLT-WINTERS de acuerdo al ANEXO 6. 72

Cuadro 8 Tabla de pronóstico por el método Holt-Winters para veintidós periodos del curso algoritmos y programación según el ANEXO 6. 74

Cuadro 9 *Tabla de resultados RMSE de los tres métodos para el curso algoritmos y programación.* 78

Cuadro 10 Tabla de pronóstico por el modelo de redes neuronales para cinco periodos del curso calculo III de acuerdo al ANEXO 5.....83

Cuadro 11 Pronóstico (2021-1 al 2028-2) por el modelo redes del curso calculo III según ANEXO 5.....84

Cuadro 12 Tabla de pronóstico para cinco periodos del curso calculo III por el método ARIMA, según ANEXO 5 90

Cuadro 13 Tabla de pronóstico por el método ARIMA para veintidós periodos del curso calculo III según ANEXO 5.....92

Cuadro 14 Tabla de pronóstico para cinco periodos para del curso calculo III por el método Holt-Winters según ANEXO 5 96

Cuadro 15 Tabla de pronóstico por el método HOLT-WINTERS para veintidós periodos del curso calculo III según ANEXO 5 97

Cuadro 16 Tabla de comparación de resultados RMSE de pronósticos del curso calculo III.....100

Cuadro 17 Tabla de matrículas en el curso de algoritmos y programación del periodo 2010-I al 2021-I.....118

Cuadro 18 Tabla de matrículas en el curso de algoritmos y programación del periodo 2010-I al 2019-II utilizados para entrenamiento.....119

Cuadro 19 Tabla de matrículas en el curso de algoritmos y programación del periodo 2020-0 al 2021-I como data de prueba y control. 120

Cuadro 20 Tabla de matrículas en el curso de cálculo III del periodo 2010-I al 2021-I120

Cuadro 21 Tabla de matrículas en el curso de cálculo III del periodo 2010-0 al 2019-II utilizados como entrenamiento.....122

Cuadro 22 Tabla de matrículas en el curso de cálculo III del periodo 2020-0 al 2021-I como data de prueba y control. 123

INDICE DE FIGURAS

Figura 1 *Valores de una Serie Temporal con Funciócn de Pronóstico y Límites de Probabilidad del 50%* 12

Figura 2 Diagrama de una Neurona Artificial 20

Figura 3 Red Perceptrón Multicapa (MLP)....22

Figura 4 Esquema Básico del Trabajo con RNA 24

Figura 5 Neurona Receptora 25

Figura 6 Ventanas Principales de la interfaz de Rstudio 36

Figura 7 Editor de texto de Rstudio....37

Figura 8 Compilador de Rstudio 38

Figura 9 Depurador en Rstudio....38

Figura 10 *Interfaz gráfica de Rstudio* 39

Figura 11 Generación de un nuevo Script....41

Figura 12 Entorno de trabajo de Rstudio 42

Figura 13 Ventanas de trabajo Rstudio....43

Figura 14 Serie temporal de matrículas del curso de algoritmos y programación 53

Figura 15 Descomposición aditiva de la serie de cursos de algoritmos y programación 54

Figura 16 Serie temporal de matrículas del curso de cálculo III. 55

Figura 17 Descomposición aditiva de la serie de cursos de cálculo III. 56

Figura 18 Análisis residual de la rede neuronal NNAR (1,1,2) [3] 58

Figura 19 Pronostico para cinco periodos del curso Algoritmos y programación por el método redes neuronales 60

.**Figura 20** *Pronostico para veintidós periodos del curso Algoritmos y programación por el método redes neuronales*.....61

Figura 21 Análisis residual por el método ARIMA 63

Figura 22 Pronostico de cinco periodos del curso Algoritmos y programación por el método ARIMA.....66

Figura 23 Pronostico de veintidós periodos del curso Algoritmos y programación por el método ARIMA.....67

Figura 24 Análisis residual por el método Holt-Winters.....69

Figura 25 Pronostico de cinco periodos por el método Holt-Winters de matrículas del curso de algoritmos y programación. 72

Figura 26 Pronostico de veintidós periodos por el método Holt- winters del curso Algoritmos y programación.....74

Figura 27 *Pronostico de los periodos (2020-0 al 2021-1) por los tres métodos para el curso algoritmos y programación.* 76

Figura 28 *Pronostico de los periodos (2021-1 al 2028-2) por los tres métodos para el curso algoritmos y programación.* 77

Figura 29 Análisis residual por Red Neuronal.....80

Figura 30 Pronostico (2020-0 al 2021-1) del curso calculo III por el método redes neuronales.....82

Figura 31 Pronostico (2021-1 al 2028-2) del curso calculo III por el método redes neuronales.....84

Figura 32 Análisis residual por el método ARIMA 86

Figura 33 Pronóstico de cinco periodos por el modelo ARIMA para el curso de cálculo III.....89

Figura 34 Pronostico de veintidós periodos por el método ARIMA para el curso de cálculo III.....91

Figura 35 *Análisis residual por el método Holt-Winters*.....93

Figura 36 Pronostico de cinco periodos (2020-0 al 2021-1) por el modelo Holt- winters para para el curso de Calculo III 95

Figura 37 Pronostico por modelo Holt- winters para veintidós periodos (2021-2 al 2028-2) para el curso de Calculo III 97

Figura 38 Pronóstico 2020-0 al 2021-1 por los tres métodos del curso cálculo III.....99

Figura 39 Pronóstico 2021-2 al 2028-2 por los tres métodos del curso de cálculo III..100

Figura 40 Pronostico para cinco periodos por Redes Neuronales calculo III tomados de Rstudio 133

Figura *41* Pronostico para veintidós periodos por Redes Neuronales para el curso calculo III tomado de Rstudio.....133

Figura 42 Pronostico para cinco periodos por el método ARIMA calculo III tomados de Rstudio 134

Figura 43 Pronostico para veintidós periodos por el método ARIMA para calculo III tomados de Rstudio.....134

Figura 44 Pronostico para cinco periodos por método Holt-Winters calculo III tomados de Rstudio.....134

Figura 45 Pronostico *para veintidós periodos por método Holt-Winters calculo III tomados de Rstudio*.....135

Figura 46 *Pronostico para cinco periodos para el curso de Algoritmos y programación por el método Redes Neuronales tomados de Rstudio*. 135

Figura 47 *Pronostico para veintidós periodos para el curso de Algoritmos y programación por el método Redes Neuronales tomados de Rstudio,* 136

Figura 48 *Pronostico para cinco periodos para el curso de Algoritmos y programación por el método ARIMA tomados de Rstudio* 136

Figura 49 *Pronostico para veintidós periodos para el curso de Algoritmos y programación por el método ARIMA tomados de Rstudio* 137

Figura 50 *Pronostico para cinco periodos para el curso de Algoritmos y programación por el método HOLT-WINTERS tomados de Rstudio* 137

Figura 51 *Pronostico para veintidós periodos para el curso de Algoritmos y programación por el método HOLT-WINTERS tomados de Rstudio* 138

CAPÍTULO I: INTRODUCCIÓN

1.1. Introducción

Es importante reconocer la naturaleza del pronóstico y sus principales aplicaciones en la economía, la bolsa de valores, el pronóstico del tiempo, de la demanda, entre otros. Su aplicación es una previsión hacia el futuro en la que se pretende acertar efectivamente a través de indicadores, estimando intervalos de error.

Los métodos de predicción se clasifican como métodos cualitativos y métodos cuantitativos, el primer método se utiliza en predicciones que no proporcionan históricos de forma directa como es el caso de nuevos productos en la cual se viene desarrollado constantemente sin un antecedente. Por lo contrario, el segundo método se basa en obtener toda la información disponible para así poder extrapolar sobre el futuro; dentro de los métodos cuantitativos se encuentra el método de descomposición, Holt-Winters, y el modelo ARIMA,

Box-Jenkis, (Uriel & Muñiz, 1993). Estos dos últimos métodos se aplican en la investigación que busca beneficiar a los stakeholders de la institución educativa con el fin de mejorar la calidad de pronóstico en la demanda de vacantes por periodo académico. Por otra parte, la investigación se alinea con la misión de la escuela profesional de Ingeniería Textil y confecciones de la UNMSM tiene como parte de su misión generar conocimiento científico para la formación de investigadores.

1.2. Planteamiento del Problema

1.2.1. Determinación del problema

En la presente investigación a tratar se sitúa dentro de la Universidad Nacional Mayor de San Marcos en Perú, escuela profesional de Ingeniería Textil y confecciones, en la cual se presentan problemas de pronóstico de demanda de alumnos al momento de la programación de secciones y profesores para la matrícula académica de los estudiantes.

Los ciclos académicos para concluir los estudios de pre grado en ingeniería están programadas para desarrollarse en 16 semanas, teniendo en cuenta que la profesión universitaria tiene una duración de 10 ciclos académicos, podremos describir una serie de problemáticas que se dan a cada inicio de ciclo en el proceso de matrícula de los estudiantes.

Al inicio de la carrera profesional, en primer ciclo, los alumnos ingresantes son asignados de forma automática a sus secciones esto se debe a la alta cantidad de estudiantes ingresantes. Los problemas de vacantes en la matricula comienzan a partir del segundo ciclo académico, según el rendimiento de cada estudiante durante el periodo de clases, se tiene como resultado casos diversos en desempeño desde alumnos que aprueban todos sus cursos hasta alumnos que desaprueban todos sus cursos, variando en la cantidad de cursos desaprobados y las veces que las desaprueban en los próximos ciclos.

Por parte de la escuela académica se realizan las programaciones de secciones para cada curso al inicio de cada ciclo teniendo en cuenta la cantidad de alumnos aprobados que vienen del ciclo anterior, repitentes, etc. las programaciones se ven reflejadas en un sistema de matrícula online en las cuales los estudiantes realizan su matrícula virtual. La matrícula virtual se

realiza en orden según el criterio: primero los estudiantes que aprobaron todos sus cursos, segundos estudiantes desaprobaron un curso al menos una vez, terceros estudiantes que desaprobaron al menos un curso dos veces. Cada uno de estos casos son programados vía online por horarios según corresponda y en distintos días, los alumnos previamente planifican posibles horarios según las secciones programadas de los cursos en los que requiere matricularse, generalmente cada curso tiene uno, dos, hasta tres secciones habilitadas en distintos horarios.

Respecto a la matricula online, los alumnos ingresan en un portal virtual donde tienen que iniciar sesión con el correo institucional y contraseña asignados a cada estudiante; se presenta una plataforma donde se carga automáticamente todos los cursos habilitados para los estudiantes, cursos en los que el estudiante ha cumplido los prerrequisitos como aprobar cursos en específico, tener aprobado una cantidad de créditos, etc.

Al momento de la matricula se visualizan los cursos, y en cada uno de estos las secciones programadas. Una vez indicado las secciones de los cursos seleccionados se procede a cargar la matrícula. Procederá a efectuarse si aún están libres las vacantes en ese instante, si es que no hay cruces de horarios en todos los cursos seleccionados.

Posterior a la matricula se evidencia que parte de ellos no llegaron a alcanzar vacantes en algunos cursos debido a la saturación del sistema, baja velocidad de internet, anulación de cursos por no llegar a la cantidad mínima de vacantes para la apertura de un curso. Esta problemática repercute directa o indirectamente en la cantidad de cursos a matricularse, disminuyendo las oportunidades al estudiante.

Al no realizar ninguna acción, los cuellos de botella por alta demanda se incrementará generando atraso académico en los estudiantes por falta de pronóstico en las secciones dejando a los estudiantes sin un docente asignado a una sección, generar reconsideraciones en los docentes programados de las secciones cerradas, generación de volumen de trabajo administrativo por la devolución de pagos, afectar en la capacidad de los cursos de laboratorio por alta demanda, dificulta la labor docente al momento de las clases y evaluación pedagógica por la cantidad de estudiantes.

Para evitar los problemas mencionados se propone implementar un sistema de pronóstico por series de tiempo debido a que no se dispone una variable independiente.

1.2.2. Formulación del problema

1.2.2.1. Problema principal

¿Cómo incide el uso del análisis de series de tiempo en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM?

1.2.2.2. Problemas secundarios

¿Cómo incide el uso de redes neuronales aplicadas a series de tiempo en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM?

¿Cómo incide la aplicación de Holt-Winters en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM?

¿Cómo incide la aplicación de modelos ARIMA en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM?

1.3. Objetivos

1.3.1. Objetivo principal

Determinar cómo influye el uso del análisis de series de tiempo en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la Universidad Nacional Mayor de San Marcos

1.3.2. Objetivos secundarios

- ✓ Determinar cómo incide el uso de redes neuronales aplicadas a series de tiempo en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.

- ✓ Determinar cómo incide la aplicación de Holt-Winters en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.

- ✓ Determinar cómo incide la aplicación de modelos ARIMA en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM

1.4. Importancia y alcance de la investigación

La investigación es importante porque busca mejorar la calidad del pronóstico de los requerimientos de vacantes en los periodos académicos por parte de los alumnos de la escuela profesional de textil y confecciones en la UNMSM

1.5. Limitaciones de la investigación

Algunas de las limitaciones de la investigación son los periodos académicos históricos en las que se desarrollaron de forma virtual durante la pandemia.

CAPÍTULO II: REVISIÓN DE LITERATURA

2.1. Marco teórico

Al estudiar los pronósticos de datos se presentan estudios previos a la investigación por el cual se busca un modelo más acertado que pueda disminuir el error de pronóstico para la toma de decisiones. Para ello es importante estudiar los conceptos de las series tiempo, sus componentes, la aplicación de modelos de resolución, los principales indicadores de control de calidad de pronóstico, entre otros. Es por ello que permite tomar mejores decisiones respecto al pronosticar datos del futuro siendo necesario abarcar sobre una base teórica.

2.2. Antecedentes del estudio

2.2.1. Internacional

Como antecedente, se puede mencionar la investigación de Botero & Cano (2008) en donde se plantea como objetivo desarrollar un método con el fin de establecer modelos de regresión en base a la data histórica de precios registrados por la bolsa de energía de Colombia, para ello se consideró una serie mensual de 144 datos, desde julio de 1995 a noviembre del 2007. Como resultado se obtuvo un modelo ARIMA el cual presentó un buen MAPE.

Otro antecedente a resaltar es el de Velasquez & Franco (2012) en donde su objetivo fue proponer una versión no lineal del modelo airline de Box y Jenkins, reemplazando el componente lineal de promedios móviles por un perceptrón multicapa. Para ello se consideró una serie de tiempo de 144 observaciones relacionada al número mensual de pasajeros en vuelos aéreos internacionales de una aerolínea entre enero de 1949 y diciembre de 1960, los resultados

obtenidos muestran que el modelo híbrido propuesto en esta investigación tiene un mejor ajuste y mejorares pronósticos que el modelo airline tradicional.

Se puede mencionar la investigación de Terán, y otros, (2019). Donde se propone desarrollar un software con el objetivo de pronosticar la demanda de asignaturas el cual se basa en el análisis de índices de aprobaciones históricas. En este sentido se desarrolla el método estadístico que se obtuvo promediando el porcentaje de aprobación y un método causal que se implementó usando red neuronal de tipo perceptrón multicapa. Para ello se consideró noventa y tres conjunto de datos de matrículas por curso en la facultad de ingeniería. La conclusión sobre el modelo causal es que produce mayores resultados regularmente y más consistente a la del modelo estadístico. No obstante, se indica en las pruebas estadísticas que no hay una diferencia significativa entre estos dos métodos.

2.2.1.1. Nacional

En la investigación de Delgado, (2020) presenta como objetivo la comparacion de tres modelos de pronostico como lo es el analisis discriminante, las redes neuronales tipo perceptron multicapa como tambien el logistico binario con el fin de evaluar el desempeño de los estudiantes del primer ciclo academico. Asi pues, el desempeño fue medido en dos factores; primero el promedio final del curso y segundo la cantidad de creditos aprobados. Asi mismo, para este fin se tomo como fuente datos de un total de 553 estudiantes correspondientes al ciclo 2017 I y 2017 II, donde los datos se agruparon en dos partes, un grupo del 70% fue de entrenamiento y el otro del 30% de validacion. Asi mismo, dentro de la aplicación de los tres modelos se utilizaron como herramientas algunos indicadores de sensibilidad como la especificidad y la curva ROC.

Por lo tanto, se concluye que es notablemente favorable la clasificacion del desempeño académico de los alumnos en los periodos 2017 I y 2017 II al aplicar modelos de regresion logistica binaria. Sin embargo, la aplicación del método de redes neuronales de tipo perceptron es más viable en la clasificacion del ciclo 2017 II siendo la mejor opcion para este periodo. Así mismo, se evidencia que en la clasificacion del desempeño academico existe un factor importante que es la nota promedio obtenido por los alumnos del modulo de matemática del examen que se rinde para la admisión.

En al investigacion de Albán, (2019). En uno de sus objetivos, se busca predecir el comportamiento de la desercion estudiantil mediante redes neuronales en la cual se desarrollan los métodos percepton multicapa y funcion de base radial. Para ello se hizo un análisis de encuesta en Google Form aplicados en distintas áreas univeritarias de ciencias técnicas, áreas administrativas como tambien humanísticas en la que se tomaron 6,000 encuestas a distintos estudiantes universitarios en el Ecuador, además la percepcion de los estudiantes sobre la deseción universitaria y sus factores busca determinar como objetivo, asi mismo la percepción de 130 expertos. La encuenta se distribuyo en tres secciones. las caracteristicas que estan relacionadas con la universidad, las secciones que estan relacionadas con los factores que estan influenciado la disercion academica y los complementos del estudio. Por lo tanto la tasa de precisión obtenida es del 96.3% y 96.8% en ambos métodos, por lo cual se concluye la fiabilidad al identificar el riesgo de la deserción temprana de estudiantes.

En la investigacion de Ancco, (2021) cuyo objetivo es encontrar el mejor modelo predictivo que se ajuste a las ventas mensuales en unidades para la jerarquizacion de calzado. Por la cual se aplicaran modelos de series de tiempo, analizando la mejor opción. Para los fines se utilizarán las ventas históricas mensuales desde el 2014 al 2021 en una primera parte se realizó un análisis

exploratorio para evaluar tendencias, estacionalidad, correlación de datos. En una sedunda parte se ajusta el modelo exponencial ETS como tambien la red neuronal autorregresiva. Finalmente comparando los errores de indicadores el modelo NNAR presenta resultados favorables.

2.3. Bases teóricas

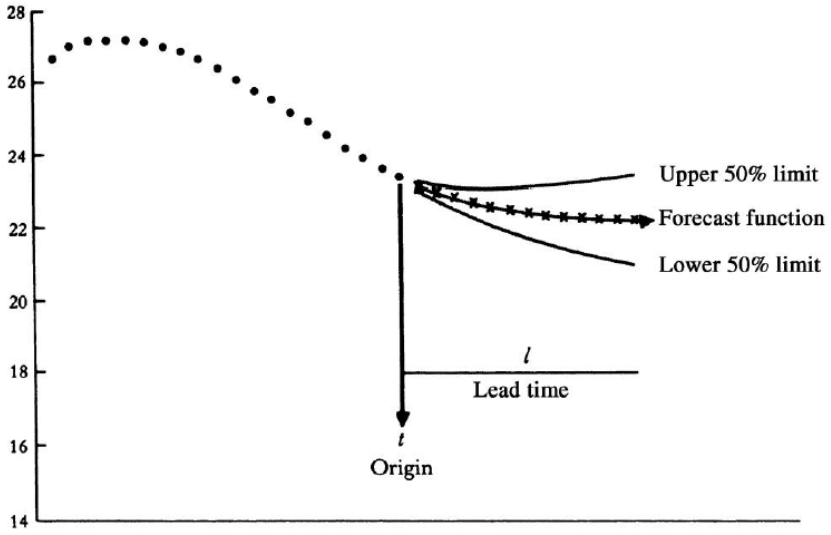
2.3.1. Series de tiempo

El estudio de las series de tiempo proporciona un evento probable en eventos futuros, un pronóstico en base a la información disponible dentro de un histórico de data. La definición vendría a darse a través de una secuencia de valores de una variable que están ordenados en intervalos simétricamente espaciados de tiempo como una herramienta de análisis, tomando como ejemplo al desempeño económico de los mercados bursátiles donde su pronóstico hace uso del histórico de la información disponible, así poder realizar un curso probable sobre su futuro. Técnicamente el pronóstico incluye análisis de los métodos univariantes, multivariante y cualitativo. Para el caso del análisis multivariable se incluye el método de regresión múltiple, análisis de la econometría como también el espacio de estado, en cambio para el análisis de metodologías cualitativas la investigación se incluye la del mercado de Delphi, el análisis situacional como también el analógico histórico; los pronósticos realizados por diversos métodos pertenecen a diferentes horizontes de tiempo. Los enfoques básicos para pronosticar se basan en extrapolar datos históricos del pasado preguntándose si los modelos sofisticados pretenden comprender el origen de los cambios del pasado e intentan construirlos dentro de sus pronósticos (Sánchez & Cuellar, 2018)

Para Box, Jenkins, Reinsel, & Ljung (2016), las series de tiempo son secuencias de observaciones tomados secuencialmente en el tiempo cuya característica intrínseca es que las observaciones adyacentes son dependientes. El análisis de esta dependencia se realiza mediante las técnicas de análisis de series de tiempo, para ello se requiere el desarrollo de modelos estocásticos y dinámicos para el tratamiento de los datos de las series de tiempo llevándolos a muchas áreas de aplicación, mediante estos métodos el análisis ayuda a tener mejores previsiones en la data, sin embargo es útil especificar una precisión de manera que se puedan calcular los riesgos al momento de tomar decisiones basadas en pronósticos, por el cual se recurre al cálculo de probabilidad convenientemente a cada lado del pronóstico. Por ejemplo, pueden ser del 50% y 95%, ver figura 1, para que cuando se produzca el valor de la serie de tiempo se encuentre dentro de este límite. Así mismo se analiza a las series multivariables como conjuntos de series individuales que componen una serie de tiempo multivariante también llamada vectorial. Para k variables en un sistema dinámico la serie estaría formado por $z_{1t}, z_{2t}, z_{3t}, \dots, z_{kt}$ para $z_t = (z_{1t}, \dots, z_{kt})'$ que denota un vector de serie temporal kx1 en tiempo t. Para describir la interrelación de las series de tiempo dentro de un vector z_t se desarrollan modelos estadísticos y análisis que permitan estudiar las relaciones dinámicas que existen entren las series de tiempo que componen el vector z_t. Como también se evidencian que difieren significativamente en la duración del tiempo en la que se consideran dentro de los métodos para pronosticar donde los métodos invariables son los que generan mayor precisión en el pronóstico a corto plazo que los multivariables.

En la siguiente figura 1 se visualiza una serie temporal con función de pronóstico y límites del 50%.

Figura 1 *Valores de una Serie Temporal con Función de Pronóstico y Límites de Probabilidad del 50%*



Nota. Para el caso se muestra un punto de origen de una serie de tiempo el cual utiliza un 50% de intervalo para que el pronóstico futuro se encuentre dentro de un intervalo convenientemente. Reproducida de Valores de una Serie Temporal con Función de Pronóstico y Límites de Probabilidad del 50%, de Box, Jenkins, Reinsel, & Ljung, 2016.

Se interpreta que es impórtate la precisión y disminución de riesgos al momento de realizar pronósticos, por esta razón se consideran probabilidades dentro de los límites del pronóstico, de esta forma asegurar el cálculo dentro de un intervalo establecido.

Según Reyes (2020), el fenómeno de variabilidad en distintos fenómenos como el clima, la temperatura, la economía, la producción es de carácter dinámico, presentan variabilidad en el tiempo; por ejemplo, el valor de las acciones dentro del mercado de valores, por momentos hay altas o bajas a tasas no constantes es decir presenta diversas irregularidades. Así pues, se en esta variabilidad se pretende encontrar un patrón regular que permita conocer el fenómeno. Dentro de sus datos históricos en cierta forma existe un patrón que se refleja en el futuro, es decir, “dentro del comportamiento irregular hay subyacente una regularidad permanente”. Así pues, se puede indicar de la serie de tiempo que se encuentra establecido por observaciones que fueron obtenidos secuencialmente en el tiempo, estas pueden ser series de forma discreta como también de forma continua. El(los) objetivo(s) de estudio de las series de tiempo pueden ser uno o varios de los siguientes: describir, predecir, explicar y/o controlar. Así mismo son el resultado de la ejecución del conjunto $\{ y_t | \, t \epsilon T\}$ donde y_t pueden ser las variables aleatorias, variables discretas como también variables continúas y t puede tomar un valor real cualquiera o en otro caso solo valores discretos: 0, ±1, ±2, ±3, ±n. En el caso que t tome valores enteros, se tratará de una serie de tiempo discreta y una serie de tiempo continua cuando los valores se encuentren un intervalo. Se conoce que, de un número finito de variables, variable discreta, se observa un solo valor Y_t^+ en cada una de las variables.

Por otro lado, se define una serie de tiempo como una relación al realizar procesos estocásticos, en la cual los procesos estocásticos son un conjunto de variables aleatorias y se puede expresar como: $\{y(t,\varsigma); t \in \Omega\}$ donde T es el conjunto índice que puede ser continuo o discreto y Ω es el espacio muestral, donde T es el tiempo. Se observa para las variables $y(t,\varsigma)$ tiene dos argumentos, un valor fijo $\varsigma = \varsigma_0$; $y(t,\varsigma_0)$ donde t es una función a realizarse posible de $\{y(t,\varsigma); t \in T; \varsigma \in \Omega\}$ y para un valor fijo de t=t_0; $y(t_0,\varsigma)$.

Por otra parte, se considera que el ruido blanco viene a ser una serie de tiempo que proviene de un proceso de variables aleatorias no correlacionadas $\{\alpha_t; t \epsilon T\}$ con $E(\alpha_t) = \mu$ y $(Var(\alpha_t) > 0$.

Como se puede apreciar las series de tiempo como observaciones de data histórica en el tiempo que tiene relación dentro de una irregularidad estas pueden ser discretas o continuas, cuyo objetivo(s) es describir, controlar, predecir, explicar.

2.3.2. Dominios de una serie de tiempo

Para Reyes (2020), las series de tiempo son estudiadas a través de dos dominios: el dominio del tiempo y el dominio de la frecuencia o análisis espectral. En el caso del dominio del tiempo, busca estudiar las variaciones de la serie en un lapso establecido a través de patrones estadísticos que vienen a ser la estacionalidad, la tendencia, los ciclos de naturaleza determinística como también una parte aleatoria los cuales fueron propuestos por Pearson. Por otra parte, el dominio de la frecuencia plantea que una serie de tiempo está conformada por las superposiciones que se dan como resultado de las ondas de las formas de senos como también cosenos que se dan a partir de las frecuencias y amplitudes diferentes donde existe un elemento aleatorio. El estudio de este comportamiento se realiza mediante un análisis espectral, este análisis toma como origen las series de Fourier.

2.3.3. Componentes de una serie de tiempo

2.3.3.1. La tendencia: $f(x)$

La tendencia según Reyes (2020), refiere como un comportamiento pre establecido por una función del tiempo $f(t) = Y_t$ donde se representa un patrón definido cuyo comportamiento puede ser lineal $Y_t = \alpha + \beta t$, parabólica: $Y_t = \alpha + \beta_1 t + \beta_2 t^2$, exponencial: $Y_t = \alpha \exp \beta t$, entre otros. El comportamiento de la serie a largo plazo está asociada a la tendencia, así mismo pueden estimarse los parámetros por métodos como el mínimos cuadrados o mínimos cuadrados ponderados teniendo como resultado métodos de suavizados exponenciales.

Como se aprecia la tendencia es un patrón definido por una función que puede ser lineal, parabólica, exponencial u otro. En la cual viene a darse dentro de una serie de tiempo a largo plazo.

2.3.3.2. La estacionalidad: $e(t)$

La estacionalidad según Reyes (2020), un fenómeno de movimientos periódicos que se repiten dentro de un intervalo de tiempo, por lo general un año, un número determinado de veces. Un ejemplo podría darse en la venta de helados que se incrementa en el verano. Dentro de la estacionalidad es posible calcular un índice de estacionalidad, como ejemplo más común se tiene al índice de Pearson, el cual es obtenido a través de medias móviles. Respecto al efecto estacional estás pueden ser aditivo como también multiplicativo, para el primer caso tiene una característica constante en cuanto al segundo caso es directamente proporcional a la tendencia por la cual se recomienda eliminar el efecto multiplicativo mediante la transformación logarítmica.

Así mismo, la estacionalidad γ_t se puede expresar mediante una adición ponderada de variables auxiliares:

$$\gamma_t = \sum_{j=1}^{s-1} Z_{j,t}\gamma_j \; ; \; \sum_{j=1}^{s-1}\gamma_j = 0 \quad , \quad Z_{j,t} = \begin{Bmatrix} 1; t = j, j+s, j+2s \ldots \\ 0; t = j, j+s, j+2s \ldots \\ -1; t = s, 2s, 3s \ldots \ldots \end{Bmatrix}$$

Donde se indica el periodo.

En consecuencia, se deduce que la estacionalidad tiene la particularidad de repetirse dentro de un determinado tiempo basado en intervalos periódicos en la cual es posible hallar el índice de estacionalidad en la cual se podrá identificar la estacionalidad y el tiempo en el que se da.

2.3.3.3. El ciclo: $c(t)$

Según Reyes (2020), el ciclo son los movimientos oscilatorios que conforman varios periodos en la cual tienen longitudes regularmente mayores a un año, en general se pueden describir en forma de funciones seno y coseno como su ajuste, en este sentido para un ciclo completo pueden interpretarse como un ciclo armónico. Así pues, podría describirse a una función del tiempo si esta es periódica, si cumple: $f(t) = f(t+p) = f(t+2p)$ ... donde p viene a ser el periodo, por lo tanto, si el desarrollo de una serie es periódico o cíclico se puede expresarse en un armónico simple como también en un armónico puro: $\gamma_t = \mu + Acos2\pi f_r(t-\varphi)$ en el cual A se refiere a la amplitud, que viene a ser la distancia entre el valor medio μ y el valor máximo de γ_t, f_r es la frecuencia la cual es el inverso del periodo $f_r = 1/p$, este valor mide los ciclos repetidos por cada unidad de tiempo, φ viene a ser la distancia del valor que toma la función valorada en su punto inicial , para t=0, y el valor del primer punto extremo que pertenece

a la curva. Así pues, toda la expresión puede ser expresada en función de la frecuencia angular. Se define w como la frecuencia angular el u al deriva de: $w = 2\pi f_r$; donde $0 \leq w \leq 2\pi$ también $\theta = 2\pi f_r\, \varphi$ por lo tanto: $\gamma_t = \mu + Acos(wt - \Theta) = \mu + Acos(wt)cos(\Theta) + Asen(wt)sen(\Theta)$, realizando los cambios $Acos(\Theta) = \alpha$, $Asen(\Theta) = \beta$ teniendo como resultado la ecuación: $\gamma_t = \mu + \alpha cos(wt) + \beta sen(wt)$. De los cambios elevando al cuadrado se tiene $A^2cos^2(\Theta) = \alpha^2$, $A^2sen^2(\Theta) = \beta^2$ el cual al sumar se tendrá: $A^2 = \alpha^2 + \beta^2$. En la expresión $\tan(\theta) = \frac{sen(\theta)}{\cos(\theta)} = \frac{\beta}{\alpha}$, $se\ obtiene\ \Theta = arctg(\frac{\beta}{\alpha})$. La serie de tiempo, el cual presenta ciclos, esta expresada como la sumatoria de armónicos simples, entonces se representa:

$$\gamma_t = \mu + \sum_{i=1}^{n} \alpha_i \cos(w_i t) + \sum_{i=1}^{n} \beta_i\, sen(w_i t)$$

Al estimar los parámetros μ , α_i y β_i se tiene en cuenta aplicar mínimos cuadrados como también las propiedades de las funciones seno y coseno considerando a una cantidad par de T observaciones, cada $w = \frac{2\pi i}{T}$; $i = 0, .., n$.

$$\alpha_i^* = (\frac{2}{T}) \sum_{i=1}^{T} y_t \cos(w_i t)\ \ para\ i = 1,2, \ldots n - 1$$

$$\alpha_i^* = (\frac{1}{T}) \sum_{i=1}^{T} y_t \cos(w_i t)\ \ para\ i = 0,1,2, \ldots n$$

$$\beta_i^* = (\frac{2}{T}) \sum_{i=1}^{T} y_t \operatorname{sen}(w_i t)\ \ para\ i = 1,2, \ldots n - 1$$

$$\beta_i^* = (\frac{1}{T}) \sum_{i=1}^{T} y_t \operatorname{sen}(w_i t)\ \ para\ i = 0,1,2, \ldots n$$

$$\mu^* = \alpha_0^*$$

Dentro de los fenómenos cíclicos se encuentran los ciclos económicos, en la cual los ciclos podrían darse dentro de periodos que se encuentran de 2 a más años por el cual se encuentra formado con variaciones de amplitud.

Según la posición se interpreta que los ciclos son movimientos oscilatorios regulares mayores a un año generalmente describen una función seno o coseno de la cual podría ser periódica describiendo una función armónica simple o de ser cíclica esta describe una función armónica pura.

2.3.3.4. El Elemento Irregular: a(t)

Según Reyes (2020), el elemento irregular se presenta como la parte irregular de una serie tras eliminar elementos estadísticos conocidos como son el periodo, la estacionalidad, entre otros. La interrelación entre estos componentes podría ser aditivo como multiplicativo según las expresiones: $\gamma_t = f_{(t)} + e_{(t)} + c_{(t)} + a_{(t)}$ para el caso aditivo y multiplicativo en los siguientes casos $\gamma_t = f_{(t)}.e_{(t)} + c_{(t)} + a_{(t)}$; $\gamma_t = f_{(t)}.e_{(t)}.c_{(t)} + a_{(t)}$. Al determinar la relación, la interrelación que guardan los elementos de una serie de tiempo entre si se desarrollaron diversas técnicas conocidas como los modelos ARIMA como también el método Box Jenkins el cual permiten entender y ampliar sobre estos modelos. Al estudiar series de tiempo implica desarrollar diversos modelos estocásticos el cual origina una serie de tiempo observado.

Se resume que el elemento irregular o aleatorio se presenta tras reducir los elementos conocidos anteriormente de una serie de tiempo por la cual tendrá que conocer una regularidad dentro de una data variable usando técnicas o modelos ARIMA.

2.3.4. Redes Neuronales

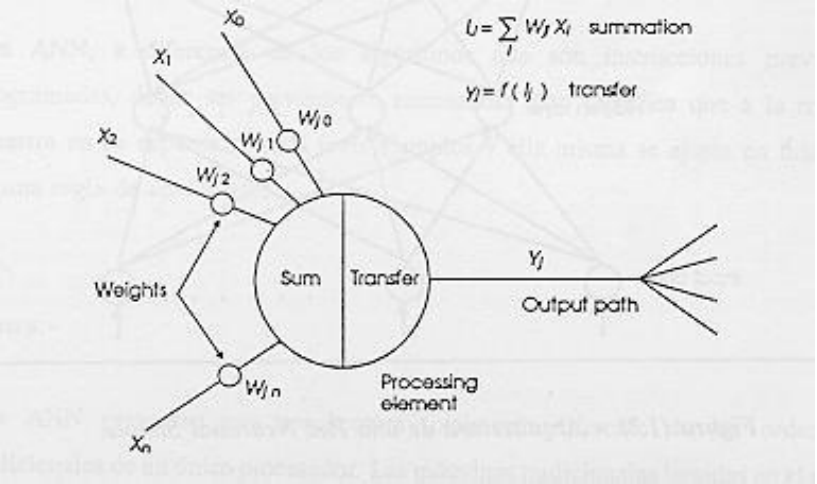
Las redes neuronales son en cierta forma parte de una inteligencia artificial. Esto debido a que se genera aprendizaje por la cual la IA sintetiza y automatiza tareas intelectuales (Callejas, Piñeros, Rocha, Hernández, & Delgado, 2013).

Como también se considera a las redes neuronales como redes neuronales biológicas que son capaces de resolver sistemas no lineales complejos altamente impredecibles, difíciles o caóticos casi imposibles de modelar (Rivas & Mazón, 2018).

Para Olabe, (1998) quien define las estructuras de una red neuronal artificial tomando como analogía un sistema nervioso humano, la neurona biológica vendría hacer el elemento procesador, PE (Process Element). Este elemento procesador posee varias entradas las cuales se combinan, la suma de las entradas es modificada mediante una función de transferencia y este valor de salida se pasa directamente a la salida del elemento procesador. Las salidas PE pueden conectarse a las entradas de otras neuronas artificiales (PE) mediante conexiones ponderadas según la ponderación sináptica, la (Figura 2) representa un elemento procesador de una red neuronal artificial.

A continuación, en la figura 2 se visualiza la red neuronal a través de un diagrama.

Figura 2 *Diagrama de una Neurona Artificial*



Nota. Las conexiones entrantes x0, …, xn son procesadas por las funciones transformadoras wj0, …, wjn teniendo como resultado yj que serán nuevos entrantes para nuevos PE. Reproducido de Diagrama de una neurona artificial (PE) de (Olabe, 1998).

Para Bishop (2006), describe a las redes neuronales asemejando a una serie de transformaciones funcionales cuyo modelo básico es $y_{(x,w)} = f(\sum_{j=1}^{M} w_j \emptyset_j(x))$ en la cual la función base viene a ser una función combinada lineal. Así mismo, los coeficientes de estas funciones combinadas linealmente actúan como parámetros adaptativos. Por otro lado, al construir M combinaciones lineales de las siguientes variables de entrada $X_1,..X_D$ en $\boldsymbol{a_j} = \sum_{i=1}^{D} \boldsymbol{w_{ji}^{(1)} x_i} + \boldsymbol{w_{j0}^{(1)}}$ donde j=1, 2…, M, el subíndice (1) refiere sobre los parámetros están en la capa de primer nivel de la red, donde se observa los parámetros $w_{ji}^{(1)}$ como pesos y $w_{j0}^{(1)}$ como sesgos, así mismo los valores de aj con conocidos como activaciones. Posterior cada uno de ellos se transforman a través de una función de activación no lineal diferenciable $h_{(aj)}$ para dar $Z_j = h_{(aj)}$ el cual corresponde a la salida de las funciones base denominadas unidades ocultas. Las funciones no lineales $h_{(aj)}$ por lo general son elegidos para ser funciones sigmoidales, estos

valores se combinan nuevamente linealmente para las activaciones de unidades de salida $\boldsymbol{a_k} = \sum_{j=1}^{M} \boldsymbol{w_{ki}^{(2)} z_j} + \boldsymbol{w_{k0}^{(2)}}$ donde k=1,….,M y k es el total de salidas, el subíndice (2) indica que los parámetros están en la capa de segundo nivel de la red, donde otra vez los parámetros de $\boldsymbol{w_{k0}^{(2)}}$ son de sesgo. En este sentido, al realizar activaciones en la unidad de salida son transformados utilizando una función de activación dando como resultado un conjunto de salidas red yk donde la decisión de elegir la función de activación va a depender de la naturaleza de los datos y posibles distribuciones de las variables objetivo. Posteriormente analiza para problemas de regresión estándar donde la función de activación es la identidad en la cual se cumple $yk = ak$. Por otro lado, en caso de problemas de clasificación binaria múltiple, la activación de cada unidad de salida se transforma a través de una función sigmoidea logística tal que $yk = \sigma(a_k)$ donde $\sigma(a) = \frac{1}{1+\exp(-a)}$,se utiliza una función de activación softmax para problemas multiclase. Combinando estas etapas, como resultado se tiene la función de red general. Para las funciones de la unidad de salida sigmoidal se obtiene la forma:

$$yk(x,w) = \sigma\left(\sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right)$$

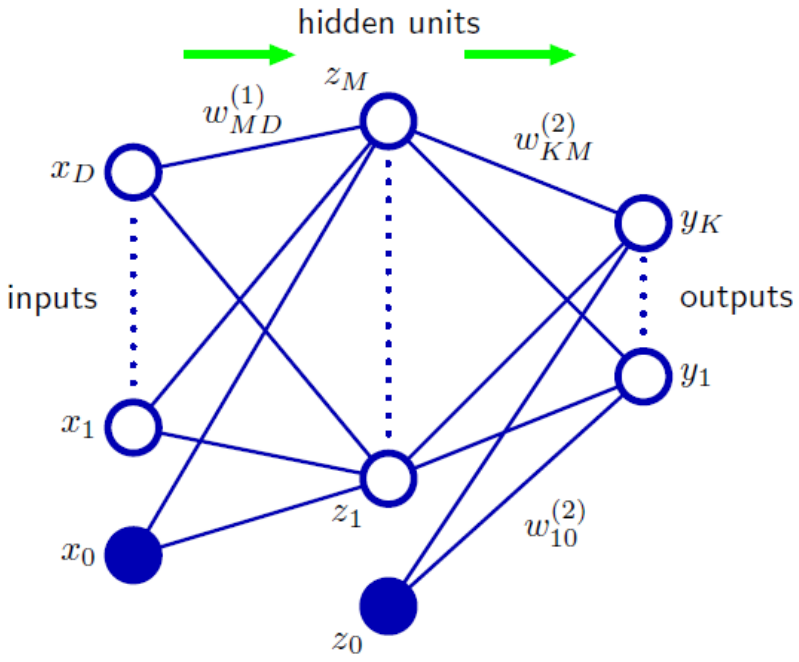
Donde todo el conjunto de parámetros de peso y sesgo se han agrupado por un vector w. En este sentido, la función se puede representar en forma de diagrama de red, ver Figura 3. Así mismo, representa a una red neuronal de dos capas en la cuales:

- ✓ Los nodos representan las variables de entrada, oculta y salida.
- ✓ Los enlaces entre los nodos representan los parámetros de peso.
- ✓ Los enlaces provenientes de nuevas entradas como de variables ocultas X0 y Z0 representan los parámetros de sesgo.

- ✓ Se puede observar que durante la propagación directa que el flujo de información toma una dirección y esta es definida por las flechas en la red.

A continuación, en la figura 3, se presenta una red perceptrón multicapa.

Figura 3 *Red Perceptrón Multicapa (MLP)*



Nota. Red neuronal de dos capas donde se consideran las variables ocultas. Reproducido de Red perceptrón multicapa (MLP), de (Bishop, 2006).

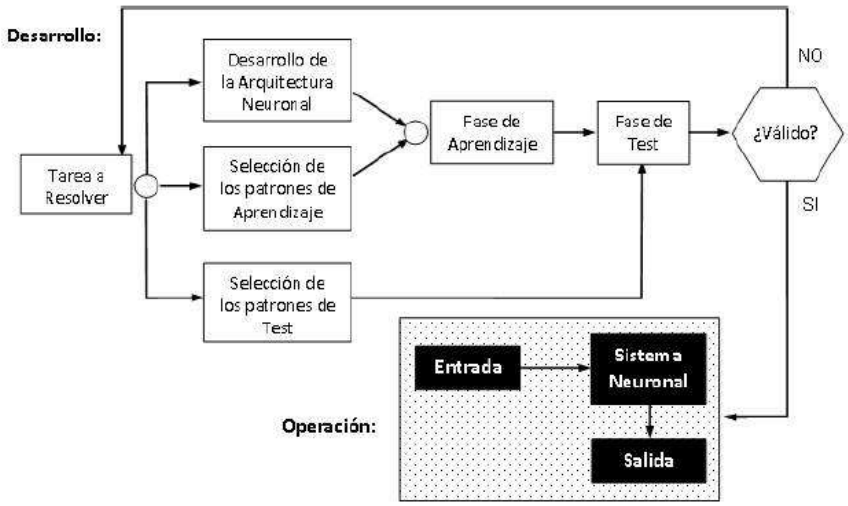
En conclusión, el modelo red neuronal está dado por una función no lineal a través de un grupo de variables que ingresan $\{x_i\}$ a la red hacia otro grupo de variables que están de salida $\{yk\}$ .Así mismo, son controlados a través del vector w cuyos parámetros son convenientemente ajustables.

2.3.4.1. Implementación de una Red Neuronal

La apreciación de Gestal (2009), sobre el uso más frecuente de las redes neuronales (RNA) por su facilidad de uso, implementación (Figura 4), donde resalta la habilidad de las redes de neuronas artificiales para hallar resultados complicados no lineales, imprecisos, resultando útil para proyectar patrones difíciles de apreciar por los humanos y técnicas computacionales. Así mismo, considera que las redes neuronales artificiales se caracterizan por tres elementos importantes: primero, la topología de la red donde se determina la capacidad del como una determinada cantidad de neuronas en capas se encuentran distribuidas entre sí y está basando en un problema planteado por resolver; segundo, la regla que permite el aprendizaje donde cierta red neuronal tiene la capacidad de aprender por entrenamiento previos debido a que es una integración de sistemas de aprendizaje; y tercero, el tipo de entrenamiento para el cual existen de dos tipos (en la etapa de aprendizaje y en la etapa de ejecución), en la etapa de aprendizaje los pesos sinápticos se adecuan en la red a través de entrenamiento y la etapa de ejecución en la cual se toman valores de funcionamiento real, esto es debió a que en esta etapa es operativa donde la red ejecuta el aprendizaje.

La figura 4 presente detalla el esquema básico de trabajo con RNA

Figura 4 *Esquema Básico del Trabajo con RNA*



Nota. Reproducido de (Gestal, 2009).

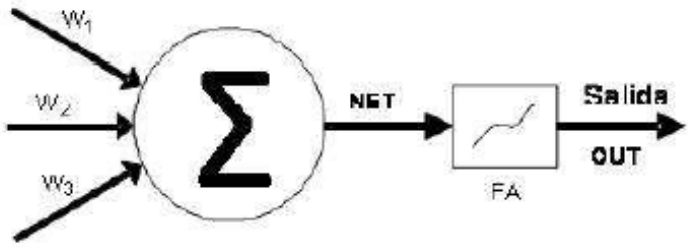
2.3.4.2. Ventajas y desventajas de una red neuronal

Algunas de las ventajas según Gestal (2009), de las redes neuronales artificiales son el aprendizaje adaptativo, la autoorganización, tolerancia a fallos, capacidad de generalización, operación en tiempo real. Tienen la capacidad de organizarse durante la etapa de aprendizaje, capacidad de retener su estructura después de daños, generalizan datos coherentes ante el ingreso de datos nuevos, el cómputo neuronal puede realizarse en paralelo. Existe una fuerza sináptica en la red neuronas el cual atribuye a un peso numérico que pondera las señales a las entradas del cual determinará la fuerza de la conexión entre dos neuronas, esta fuerza sináptica se considerara alto o una sinapsis fuerte entre dos elementos cuando la contribución de la información es a gran escala o altera notoriamente a la neurona receptora, por los tanto dependerá de esta recepción la respuesta que esta transmita. Para la evaluación de una neurona se calcula todas las fuerzas de entradas (NET), posterior se utiliza la función de activación (FA)

para determinar internamente el valor de una neurona la cual será el valor que transmita a su salida. Las combinaciones que recibe una neurona se pueden determinar por $NET_i(t) = \sum_{j=1}^{n-1}[W_{ij}.O_j.(t-1)]$, donde W_{ij} viene a ser el peso de una conexión entre la neurona emisora j y la neurona receptora i.

En la presente figura 5 se visualiza a la neurona receptora.

Figura 5 *Neurona Receptora*



Nota. Las funciones de transferencia y activación se encargan de definir un estado nuevo de activación Ai y Oi como respuesta de la neurona. Así mismo, las entradas que recibe la neurona en un periodo dado como también el estado de activación que tenga son dependientes de acuerdo al estado de activación como la salida de la red. Reproducido de (Gestal, 2009).

Para Rivera (20015, p. 64), la desventaja de la red neuronal construida artificialmente se encuentra en la realización de las tareas a gran escala. Así mismo, tienen elevado tiempo de aprendizaje puesto que dependerá del número de patrones en la serie. En este sentido depende de la cantidad enorme de ejemplos que se necesita para el entrenamiento, entre otros.

2.3.5. ARIMA

Según Sánchez & Cuellar (2018), el objetivo del modelo ARIMA es proporcionar herramientas útiles para extrapolar de una forma segura información de relevancia basada en datos que forman series de tiempo. El enfoque es la implementación de modelos predictivos univariables o de una sola serie a la en la cual la transferencia polinómica racional es la principal herramienta de pronóstico y menudo se presentan residuos de modelamiento en las series de tiempo, esto debido al momento de la operación de pronóstico se presentan situaciones. Así mismo, los datos de las variables independientes pueden no estar disponibles o no ser de entera confianza, esta serie de errores pueden presentar variación o una correlación con el fin de buscar en los residuos menorar su correlación. Además, las funciones de transferencia pueden ampliarse y refinarse a través de un Auto-Regresivo (AR) o promedio móvil (MA).

En conclusión, el propósito es desarrollar una base en la cual se puedan tomar decisiones al comparar modelos y enfoques alternativos o hacer adecuaciones de modelos como en la presente investigación para describir el comportamiento de la congestión de redes de datos.

Modelos matemáticos ARIMA

Para Sánchez & Cuellar (2018), Los modelos matemáticos ARIMA que los compone son:

2.3.5.1. Modelos Autorregresivos (AR)

Son los modelos que describe determinados procesos en la cual las observaciones en un determinado momento se pueden predecir tomando en cuenta a las observaciones previas como también la identificación un término de error de la serie. Caso ARIMA (1,0,0) O AR (1).

$$Y_t =_i Y_{t-1} + a - t$$

$$Z_t = \sum_{j=1}^{\infty} j\, Z_{(t-j)} + a_t$$

2.3.5.2. Modelo de promedios móviles (MA)

Es el modelo que describe una serie estacionaria de tiempo, para este caso en el modelo se podría predecir el valor actual a partir del de los impulsos aleatorios registrados históricamente en el tiempo. CASO ARIMA (0,0,1) o MA (1).

$$Y_t = a_t - \theta a_{t-i}$$

$$Z_t = \mu + \sum_{j=1}^{\infty} j\, a_{(t-j)}$$

2.3.5.3. ARIMA(p,d,q)

Este modelo busca describir a un valor como una función lineal de datos históricos y a los errores surgidos aleatoriamente, al azar. Este análisis se realiza en una serie que presenta estacionalidad, para al menos 50 datos.

Para Akiva, (1955), el modelo ARIMA es una relación combinatoria de un término autorregresivo (AR) y un término de promedio móvil (MA) con un término diferenciador (I).

2.3.5.4. Proceso autorregresivo $AR_{(p)}$

En este proceso cierta serie de tiempo $X_{\{t\}}$ cumple con un proceso autorregresivo $AR_{(p)}$ cuyo orden será p si se cumple:

$$X_{\{t\}} = \emptyset_1 x_{\{t-1\}} + \cdots + \emptyset_p x_{\{t-p\}} + Z_{\{t\}}$$

Donde $\emptyset_1, \ldots., \emptyset_p$ son constantes y $Z_{\{t\}} \sim N_{(0,\sigma^2)}$ conocido como ruido.

2.3.5.5. Proceso de promedio móvil $MA_{(q)}$

Es el proceso donde cierta serie de temporal $X_{\{t\}}$ cumple como un proceso móvil $MA_{(q)}$ cuyo orden es q si cumple:

$$X_{\{t\}} = Z_{\{t\}} + \theta_1 Z_{\{t-1\}} + \cdots + \theta_q Z_{\{t-q\}}$$

Donde $Z_{\{t\}} \sim N_{(0,\sigma^2)}$ y $\theta_1, \ldots., \theta_q$ son constantes

Por lo general, estos modelos refieren a ARIMA(p,d,q)

- ✓ Donde p está vinculado al orden del modelo autorregresivo.
- ✓ Donde d está vinculado a la diferenciación.

- ✓ Donde q termino de media móvil en q elementos de error.

2.3.5.6. Proceso de diferenciación I(d)

En este proceso se sugiere que el termino referencial debe evaluar el orden, donde el grado del diferencial podría ser de primero, segundo o grado d. considerando la forma siguiente que tienen las ecuaciones:

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

En el caso de B se definirá como operador de retraso ($Bx_{\{t\}} = x_{\{t-1\}}$)

La diferenciación de segundo orden será:

$$\nabla^2 x_{\{t\}} = (1 - B)^2 x_{\{t\}} = (1 - B)(1 + B)x_{\{t\}} = (1 - 2B + B^2)x_{\{t\}} = x_{\{t\}} - 2Bx_{\{t\}} + B^2 x_{\{t\}}$$

Para lo cual, la diferenciación de orden d se definirá como:

$$\nabla^d x_{\{t\}} = (1 - B)^d x_{\{t\}}$$

Por lo tanto, la estructura general de estos modelos $\{Y_{(t)}; t \epsilon T\}$ en la cual T representa a los instantes de tiempo de cierta serie y se representa como la ecuación:

$$\hat{y}_{\{t\}} = \sum_{i=1}^{p} \emptyset_i Y_{\{t-i\}} - \sum_{j=1}^{q} \theta_j Z_{\{t-j\}}$$

- ✓ $\emptyset$ corresponde al coeficiente autorregresivo a determinar
- ✓ θ corresponde al coeficiente de media móvil a determinar
- ✓ z corresponde al termino de error de la serie a modelar
- ✓ $Y_{\{t\}}$ es el registro normalizado de la serie a modelar

El objetivo del modelo es encontrar parámetros para $\emptyset\ y\ \theta$ tal que se posible modelar un comportamiento futuro para estimar $\hat{y}_{\{t\}}$ a partir de valores anteriores $y_{\{t-1\}}, \ldots., y_{\{t-p\}}$ como también $Z_{\{t-1\}}, \ldots., Z_{\{t-q\}}$, se considera valido el modelo siempre en cuando se cumpla:

$$\emptyset(B)(1-B)^d X_{\{t\}} = \theta(B) Z_{\{t\}}$$

- ✓ $\emptyset$ es un polinomio de grado p
- ✓ θ es un polinomio de grado q

2.3.6. Holt-Winters

Para Box, Jenkins, Reinsel, & Ljung(2016), los componentes que permitan realizar pronósticos a través del potenciamiento del método del suavizamiento exponencial simple son la tendencia, estacionalidad, el nivel base como también el componente ruido. Estas podrían ser del tipo aditivo como el del multiplicativo, esto dependerá de la constancia o cambio de la estacionalidad, estas se detallan a continuación.

Aditivo, se basa en la constancia en el tiempo de la estacionalidad en el transcurso del tiempo.

$$\hat{y}_{t+h} = L_t + hT_t + S_{t+h-m\left[\frac{h-1}{m}\right]}$$

Multiplicativo, se basa en el caso que la estacionalidad se comporte proporcionalmente al nivel de la serie.

$$\hat{y}_{t+h} = (L_t + hT_t) \times S_{t+h-m\left[\frac{h-1}{m}\right]}$$

Asimismo, desarrollar una técnica que permita llevar a la suavización exponencial simple a una suavización exponencial lineal será mediante el manejo de datos con tendencias puesto que permite incluir componentes de tendencia y estacionalidad. Este método de pronóstico de suavizamiento exponencial deberá utilizar tres parámetros de suavizamiento α, β , γ (valores entre 0 y 1) como también su relación para le pronostico según los siguientes componentes.

Nivel de suavización $(\boldsymbol{L_t})$, permite ponderar la observación actual y_t con α como también la suma ponderada del nivel anterior L_{t-1} más la tendencia anterior T_{t-1} con $\alpha - 1$d con esto se refleja la importancia relativa de la observación actual con el nivel/tendencia anterior en la estimación del nivel actual.

Aditivo

$$L_t = a(y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1})$$

Multiplicativo

$$L_t = a(y_t \,/\, S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1})$$

Tendencia de suavización $(\boldsymbol{T_t})$, esta ecuación modela la suavización de la ecuación donde se ponderan con β la diferencia de los niveles actuales y anteriores, asimismo se suma con $(1-\beta)T_{t-1}$ de esta forma se actualiza la información sobre la tendencia.

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

Estacionalidad ($\boldsymbol{S_t}$), permite estimar los periodos repetitivos, Patrón cíclico, que ocurre cada m periodos.

Aditivo

$$S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-m}$$

Multiplicativo

$$S_t = \gamma\left(\frac{y_t}{L_t}\right) + (1 - \gamma)S_{t-m}$$

Según Ezequiel J., Manuel. M. (1993), observan que otra alternativa para analizar en general las series de tiempo puede ser a través de los métodos típicos; se basa en la descomposición de la serie. Para todos los casos una serie puede ser descompuesta en una o varios elementos como la tendencia, el factor cíclico, la estacionalidad, el componente irregular. Así pues, el método de descomposición plantea que las series de tiempo representan a los resultados al integrar todos estos elementos de forma aditiva (la tendencia no afecta las fluctuaciones) o de manera multiplicativa (la tendencia reproduce cambios en las fluctuaciones).

En este sentido, si una serie está representada en un esquema multiplicativo y a su vez tiene estacionalidad, el método apropiado selecto sería el de media móvil por tener consistencia y propiedad de eliminar los factores estacionales. Así pues, al desestacionalizarse la serie se encuentra en la capacidad de realizar predicciones más confiables para el futuro.

Por otro lado, las series que presentan una tendencia aproximada linealmente y que son sometidas a una incidencia del factor estacional generalmente se recomienda aplicar el método Holt-Winters. La aplicación de este método servirá de base para la predicción expresada de la siguiente forma: $Y_t = (b_0 + b_1)E_t + \mu_t$ donde el factor b_0 es permanente, b_1 es la pendiente de una recta y el factor E_t es estacional multiplicativo.

A continuación, para estimar los componentes de este método se plantean tres ecuaciones:

$$S_t = \alpha \frac{Y_t}{C_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad 0 < \alpha < 1$$

$$b_{1t} = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1} \qquad 0 < \beta < 1$$

$$C_t = \gamma \frac{Y_t}{S_t} + (1-\gamma)C_{t-L} \qquad 0 < \gamma < 1$$

En conclusión, para aplicar el método Holt-Winters al realizar predicciones será necesario conocer los valores iniciales y los valores de las constantes α, β, γ. Para realizar algunos cálculos recursivos lo valores iniciales serán L+2, según corresponda a los L factores con estacionalidad del periodo pasado, a la pendiente y nivel del periodo 0, a la primera observación.

2.3.7. Lenguaje de programación R

R fue creado por Ross Ihaka y Robert Gentleman, en la cual el fin era crear un lenguaje de introducción a la estadística didáctico para ser dictado como curso en la universidad de nueva Zelanda. Por la cual se decidió usar como base la sintaxis similar al lenguaje S desarrollado por Bell Laboratories. Sin embargo, la semántica es sencillamente sensiblemente diferente (Ihaka, 1998).

Así mismo, Gough B. (2009), precisa que desde 1995 la fuente del código de R está a disposición bajo la licencia publica GNU en los sistemas operativos Windows, MacsOS y Linux. La licencia se presenta de forma libre bajo la fundación R (R Fundation), en el cual se puede estudiar, modificar, compartir, copiar, mejorar sin ninguna restricción.

Para Santana & Farfán, (2014), R se encuentra interpretado en un ambiente para el computo gráfico y estadístico. El programador podrá comprender organizadamente instrucciones de forma que le permita analizar serie de datos para la mejora en la toma de decisiones. Así mismo, mencionan que R pose los conceptos y características siguientes.

Conceptos:

1) El sistema base de R, se puede bajar de CRAN (Comprehensive R Archive Network).
2) La funcionalidad de R consta de paquetes modulares. R contiene el paquete básico para su sistema base que se requiere para su ejecución como también la mayoría de las funciones fundamentales.

Características por ser gratuito:

1) Ejecútalo para cualquier propósito.
2) Estudiarlo y adaptarlo para cualquier propósito.
3) Redistribuir copias para distintos propósitos.
4) Posibilidad de mejorar las versiones.

Las principales ventajas que se describes para R es que es gratuito de uso aplicable en multiplataforma, todos los sistemas operativos. Desde R se pueden leer con facilidad archivos generados desde otros programas como SAS, Stata, Minitab, Excel, entro otros. Así mismo, una de las principales desventajas se sitúa en el laborioso trabajo de codificar líneas a la cual se da paso a Rstudio. Teniendo una interfaz más amigable fácil de depurar, organización y optimización del espacio de trabajo. Así mismo, Rstudio se encuentra en ediciones comerciales cuyo código

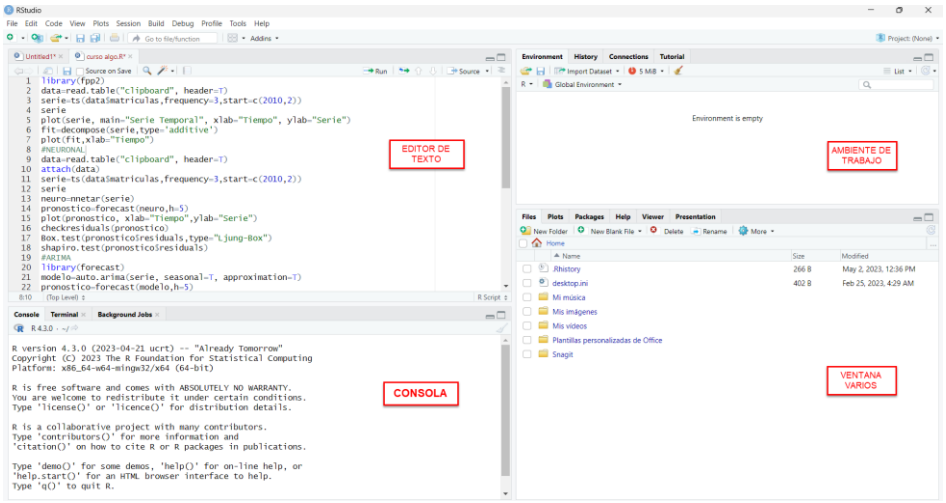
es abierto en la que se puede ejecutar en cualquier escritorio como Windows, Mac o Linux, (Rstudio, 2023).

2.3.8. Rstudio

Rstudio es un entorno de desarrollo integrado (IDE) que se usa para R y Python, se desarrolla mediante una consola donde existe un editor de sintaxis resaltable la cual permite una ejecución precisa de código como también herramientas para las gráficas, la gestión de (Rstudio,2023)

Para Vargas L., Fuquen E. (2021), el IDE es un programa compuesto por cuatro herramientas y dos áreas principales. En este sentido, se puede interpretar el lenguaje R mediante Rstudio visto como una interfaz capaz de interactuar para poder “comunicarse” con el computador. De igual modo, la interfaz está compuesta por cuatro ventanas principales: el editor de textos o script, la consola, el entorno de trabajo y “ventana varios”, ver Figura 6.

Figura 6 Ventanas Principales de la interfaz de Rstudio



Fuente: Elaboración propia

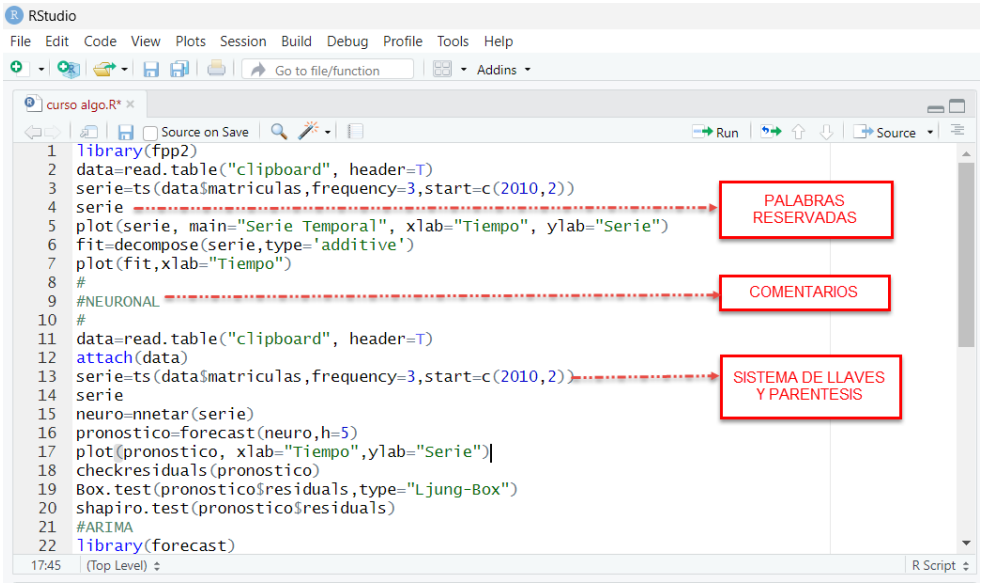
Herramientas de IDE

A continuación, se presentan las cuatro herramientas de IDE. Rstudio para la presente investigación en la cual se detallarán las funcionalidades, características de cada uno de ellos.

1) Editor de texto

Permite escribir instrucciones de forma organizada, sin tener ningún tipo de formato como son los formatos cursiva, negrilla, texto centrado, entre otros.

Figura 7 *Editor de texto de Rstudio*

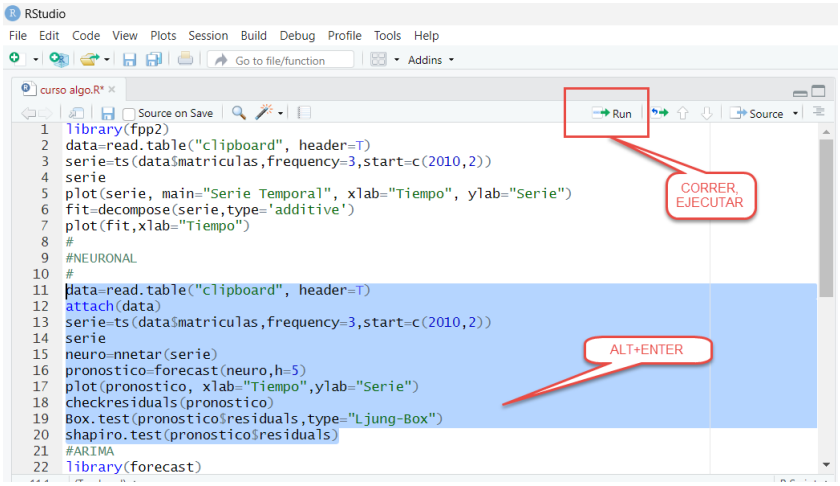


Fuente: Elaboración propia

2) Compilador

Permite traducir el código introducido en R, es el encargado de interpretar las instrucciones para que pueda ser ejecutado mediante un comando o selección del código y Alt+Enter (ver Figura 8). Es decir, en primera instancia valida la sintaxis del código máquina que esté correctamente para luego devolver un resultado.

Figura 8 *Compilador de Rstudio*

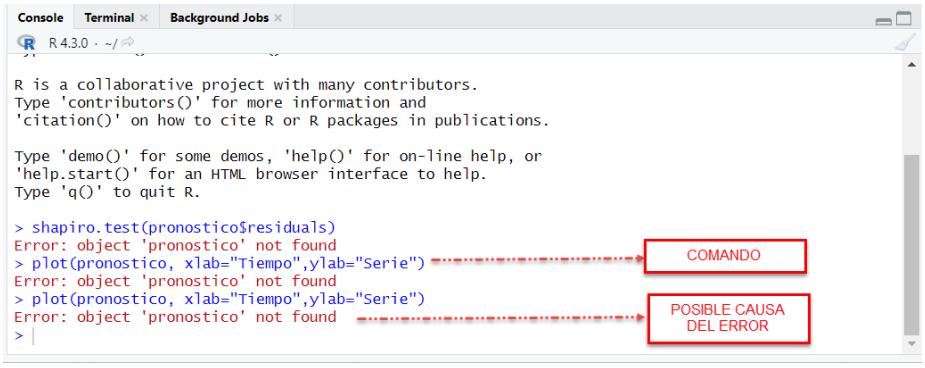


Fuente: Elaboración propia

3) Depurador

Es una herramienta que permite visualizar e identificar los errores en el código, es importante mencionar que una vez se detecte el error, el depurador expone los errores, las posiciones y posibles causas.

Figura 9 *Depurador en Rstudio*

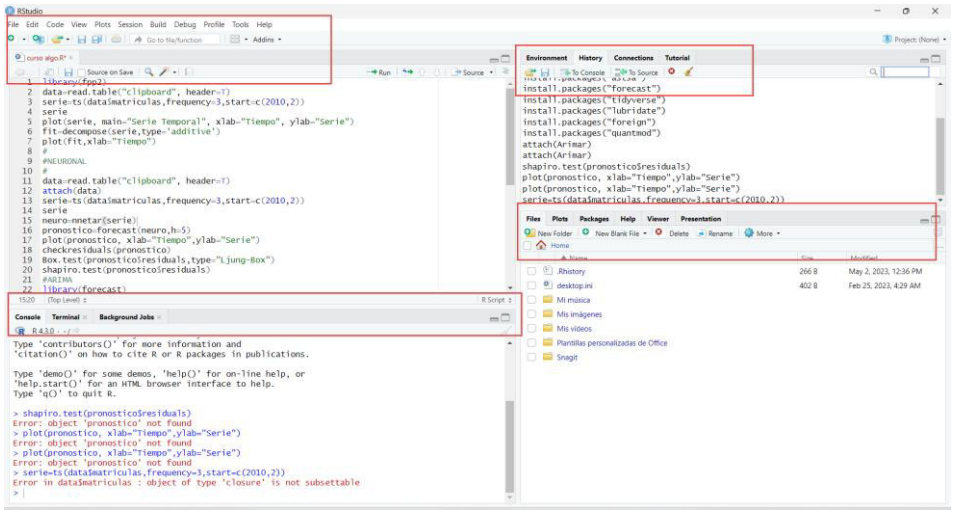


Fuente: Elaboración propia

4) Interfaz gráfica

Presenta un entorno amigable la cual permite tener comodidad y versatilidad, puesto a que el usuario interactúa de manera dinámica con el computador con el manejo de la programación en R.

Figura 10 *Interfaz gráfica de Rstudio*



Fuente: Elaboración propia

Áreas principales

A continuación, se presentan las áreas principales conformado por un área de entrada (consola y script), como también un área de salida (entorno de trabajo, ventana varios).

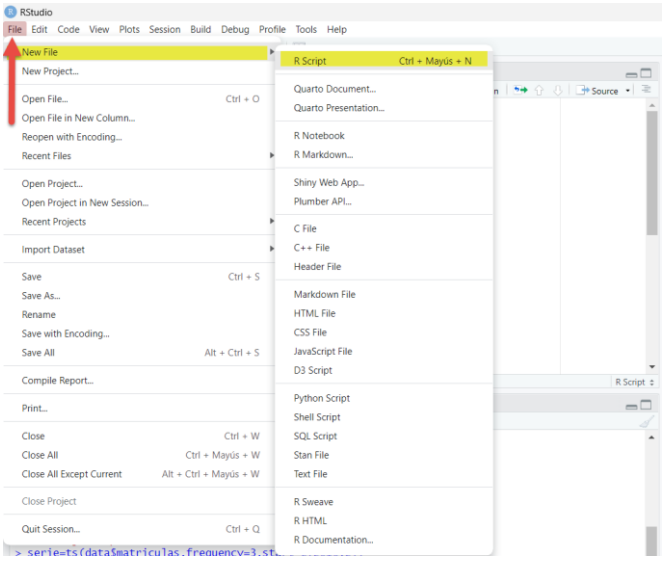
1) Área de entrada

Está compuesta por las ventanas de la consola y el script, en estas ventanas se cumplen la tarea de ingresar instrucciones que posteriormente serán ejecutadas.

I. Consola. - se visualiza en la parte izquierda del programa, entre sus características están que se muestran los códigos ingresados donde se muestran ya validados con sus respectivas observaciones o errores (alertas) obtenidos al ejecutar el código. A todo esto, es una ventana donde se ofrece una comunicación entre el software y el usuario.

II. Script. – Es el editor de texto que permite plasmar, ejecutar el código. El script, a diferencia de la consola, permite guardar los códigos para usarlos en próximas codificaciones. Para poder crear un nuevo script es necesario abrir en la parte izquierda superior en archivos, nuevo archivo, R script, una vez abierto se apertura un script nuevo parecido a un formato bloc de notas, en la cual se ofrecen diversas opciones de guardado como también para ejecutar el código.

Figura 11 *Generación de un nuevo Script*



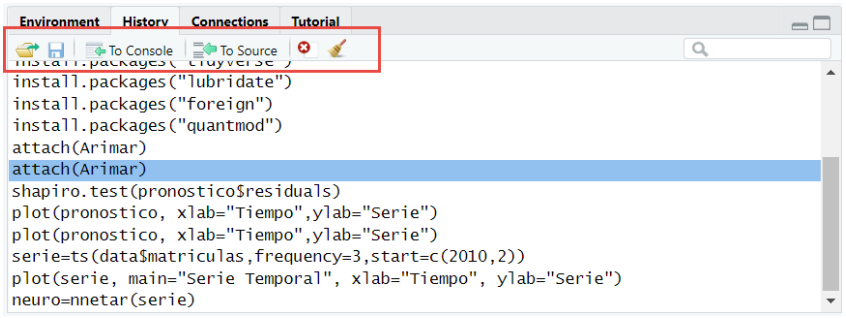
Fuente: Elaboración propia

2) Área de salida

Esta área presenta tres partes: el ambiente de trabajo, la ventana varios y la consola las cuales se encargan de mostrar resultados a través de gráficos, ofrecer ayuda con manuales para facilitar el manejo de R como también Rstudio.

I. Ambiente o entorno de trabajo. – se encuentra en la parte derecha de la interfaz, presenta las funciones, variables, constantes usados con sus resultados obtenidos al ejecutar los códigos. Es recomendable usar una extensión, Rdata, para salvaguardar el archivo del ambiente de trabajo.

Figura 12 *Entorno de trabajo de Rstudio*

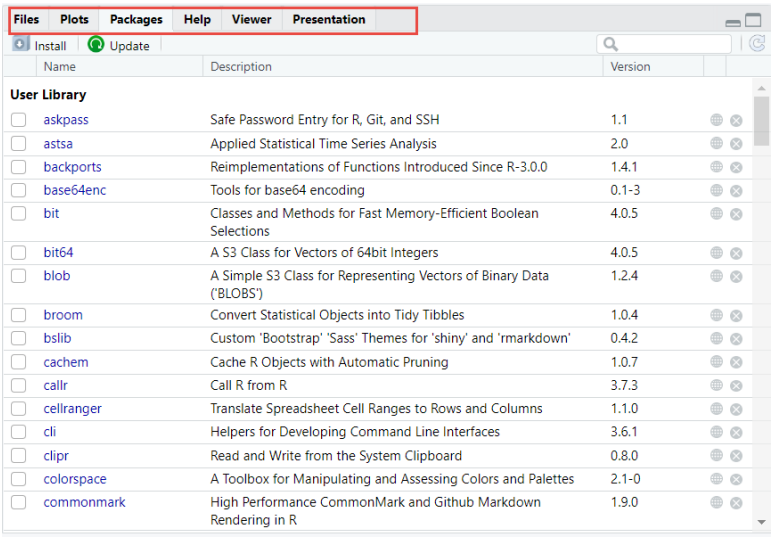


Fuente: Elaboración propia

II. Ventana varios. – está ubicado en la parte inferior derecha de la interfaz donde se encuentran cinco fichas de trabajo (figura N), que se muestran a continuación:

- Files (explorador de archivos)
- Plots (visualización de gráficos)
- Packages (Paquetes de R)
- Help (documentación de ayuda Rstudio)
- Viwer (visualizaciones de datos)
- Presetation (formatos de presentación)

Figura 13 *Ventanas de trabajo Rstudio*



Fuente: Elaboración propia

2.4. Definición de términos

Inteligencia Artificial (IA): Es la habilidad que tienen las máquinas para usar algoritmos, aprender de los datos y usarlos para tomar decisiones como lo harían los humanos (Rouhiainen, 2018)

Redes Neuronales Artificiales (RNA): Es parte de la inteligencia artificial puesto que genera aprendizaje, evalúa y toma decisiones (Callejas, Piñeros, Rocha, Hernández, & Delgado, 2013).

Tendencia estadística: viene hacer el comportamiento o la manera de comportarse de un sujeto, opción, muestra o universo (García, 2011).

Estimación: “juicio sobre el valor del resultado de una operación numérica o de la medida de una cantidad, en función de circunstancias individuales del que lo emite” (Castro, Castro, Rico, & Segovia, 1989).

Sinapsis: La sinapsis puede ser definida como aquellas uniones especializadas en las cuales una neurona se comunica con una célula diana (Nava, 2017, pp. 85-92)

Pronóstico: “es una estimación cuantitativa o cualitativa de uno o varios factores (variables) que conforman un evento futuro, con base en información actual o del pasado” (Villareal, 2016, pp. 1-121).

Perceptrón: “Un perceptrón es una neurona artificial, y, por tanto, una unidad de red neuronal. El perceptrón efectúa cálculos para detectar características o tendencias en los datos de entrada” (Rosenblatt, 1957).

CAPÍTULO III: HIPÓTESIS Y VARIABLES

3.1. Hipótesis

3.1.1. Hipótesis principal

El uso del análisis de series de tiempo incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.

3.1.2. Hipótesis secundarias

- ✓ El uso de redes neuronales aplicadas a series de tiempo incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y confecciones de la UNMSM

- ✓ La aplicación de Holt-Winters incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM

- ✓ La aplicación de modelos ARIMA incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.

3.2. Variables

HG: El uso del análisis de series de tiempo incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.

Variable Independiente: series de tiempo

Variable Dependientes: Calidad, Pronostico de la demanda

HE1: El uso de redes neuronales aplicadas a series de tiempo índice en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y confecciones de la UNMSM

Variable Independiente: Redes Neuronales

Variable Dependiente: Pronostico de la demanda

HE2: La aplicación de Holt-Winters incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM

Variable Independiente: Holt-Winters

Variable Dependiente: Pronostico de la demanda

HE3: La aplicación de modelos ARIMA incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.

Variable Independiente: Modelos ARIMA

Variable Dependiente: Pronostico de la demanda.

3.4. Operacionalización de las variables

En el siguiente cuadro 1 se analiza la operacionalización de las variables

Cuadro 1 *Operacionalización de las variables*

"Influencia de la aplicación de técnicas econométricas de series de tiempo y Machine Learning en el pronóstico de la demanda de cursos de la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM"			
Variables Independientes	**Definición conceptual**	**Definición Operacional**	**Indicadores**
Series de tiempo	Son secuencias de observaciones tomadas secuencialmente en el tiempo cuya característica es que las observaciones adyacentes son dependientes (Box, Jenkins, Reinsel, & Ljung, 2016).	Permite conocer la data histórica con el objetivo de hallar los posibles pronósticos.	Se Aplica / No se aplica
Redes Neuronales	Son parte de la inteligencia artificial debido a que genera aprendizaje mediante el cual sintetiza y automatiza tareas intelectuales (Callejas, Piñeros, Rocha, Hernández, & Delgado, 2013).	Es la parte operativa por la cual se pueden diseñar métodos de extrapolación de información.	Se Aplica / No se aplica
Holt-Winters	El método Holt-Winters es otra alternativa para analizar series temporales, también llamado métodos de descomposición debido a sus cuatro componentes: tendencia, factor cíclico, estacionalidad, componente irregular (Uriel & Muñiz, 1993)	Es un método a pivotear en el cual se podrá buscar la mayor calidad de pronóstico.	Se Aplica / No se aplica
ARIMA	Los métodos ARIMA se basan en proporcionar herramientas para extrapolar información de forma segura basado en datos de series de tiempo (Sánchez & Cuellar, 2018).	Permite obtener otra posibilidad de pronóstico en el tratamiento de series de datos.	Se Aplica / No se aplica

Variable Dependiente	Definición conceptual	Definición Operacional	Indicadores
Calidad del pronóstico de la demanda	Al realizar el uso del pronóstico implica realizar el tratamiento de información accesible para realizar cálculos probables sobre la orientación de los eventos a futuro, para el caso la demanda de alumnos (Sánchez & Cuellar, 2018).	Es indispensable ante cualquier situación contextual que surge dentro de cierta situación de mejora.	ME, RMSE MAE

Fuente: Elaboración propia

3.5. Matriz de consistencia

En el siguiente cuadro 2 se presenta la relación entre el problema, objetivo, hipótesis, variables e indicadores.

Cuadro 2 *Matriz de consistencia*

"Influencia de la aplicación de técnicas econométricas de series de tiempo y Machine Learning en el pronóstico de la demanda de cursos de la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM"				
Problema General	**Objetivo General**	**Hipótesis General**	**Variables**	**Indicadores**
¿Cómo incide el uso del análisis de series de tiempo en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM?	Determinar cómo influye el uso del análisis de series de tiempo en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.	El uso del análisis de series de tiempo incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.	Variable Independiente: -series de tiempo Variables Dependientes: - Calidad del pronóstico de la demanda	- Se Aplica / No se Aplica -ME, RMSE MAE

Problemas Específicos	**Objetivos Específicos**	**Hipótesis Especificas**	**Variables**	**Indicadores**
¿Cómo incide el uso de redes neuronales aplicadas a series de tiempo en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM?	Determinar cómo incide el uso de redes neuronales aplicadas a series de tiempo en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.	El uso de redes neuronales aplicadas a series de tiempo índice en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y confecciones de la UNMSM	Variable Independiente: - Redes Neuronales Variables Dependiente: - Calidad del pronóstico de la demanda	- Se Aplica / No se Aplica -ME, RMSE MAE
¿Cómo incide la aplicación de Holt-Winters en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM?	Determinar cómo incide la aplicación de Holt-Winters en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.	La aplicación de Holt-Winters incide en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.	Variable Independiente: - Holt-Winters Variables Dependiente: - Calidad del pronóstico de la demanda	- Se Aplica / No se Aplica -ME, RMSE MAE
¿Cómo incide la aplicación de modelos ARIMA en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM?	Determinar cómo incide la aplicación de modelos ARIMA en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM	La aplicación de modelos ARIMA incide en el pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM.	Variable Independiente: - Modelos ARIMA Variables Dependiente: - Calidad del pronóstico de la demanda	- Se Aplica / No se Aplica -ME, RMSE MAE

Fuente: Elaboración propia

CAPÍTULO IV: MATERIALES Y MÉTODOS

4.1. Área de estudio

El presente trabajo se encuentra dentro del objetivo de investigación de educación de calidad y línea inteligencia artificial debido a que se usaran pronósticos en base a las redes neuronales, la cual tiene amplias aplicaciones como en el caso de la presente investigación.

4.2. Tipo y diseño de investigación

El presente proyecto se enmarca en el tipo Descriptivo Explicativo. El diseño se ilustra en el siguiente esquema (Sánchez y Reyes, 2009):

M O 1

En donde:

O: Observación o levantamiento de datos de la serie

M: Muestra o serie de tiempo

4.3. Población, muestra y unidad de análisis

Población

Finita, todas las matrículas de alumnos en los cursos de algoritmos de los periodos 2010 I al 2021 I. y programación y cálculo III del periodo 2010 0 al 2021 I.

Muestra

Entrenamiento. - matriculas de los alumnos en los cursos de algoritmos y programación de los periodos 2010 I al 2019 II y cálculo III de los periodos 2010 0 al 2019 II.

Prueba y control. - Matriculas de los alumnos en los cursos de algoritmos y programación y cálculo III de los periodos 2020 0 al 2021 I

Muestreo

Probabilístico por conglomerados.

Unidad de análisis

Matrículas de alumnos en los cursos de algoritmos y programación y cálculo III

4.4. Procedimientos, técnicas e instrumentos de recolección de datos

La técnica usada en la investigación de recolección de datos fue la revisión de base de datos a través de los reportes del SUM (sistema único de matrícula) de los alumnos matriculados en los cursos de algoritmos y programación en los periodos 2010 I al 2021 I y calculo III en los periodos 2010 0 al 2021 I.

4.5. Análisis estadístico

De acuerdo al análisis de datos se formuló el modelo de pronóstico aplicando redes neuronales, el método ARIMA y Holt Winters. Del mismo modo, para la validación de estos

modelos se deberán probar los tres estadísticos de control para validar su viabilidad (los residuales deben estar dentro del intervalo de confianza ACF, siguen una distribución normal mediante el test Shapiro-Wilks y se distribuyen de forma dependiente mediante el test Ljung-Box). Asimismo, para la medición de la calidad del pronóstico se utilizó el estadístico de error RMSE.

CAPÍTULO V: RESULTADOS

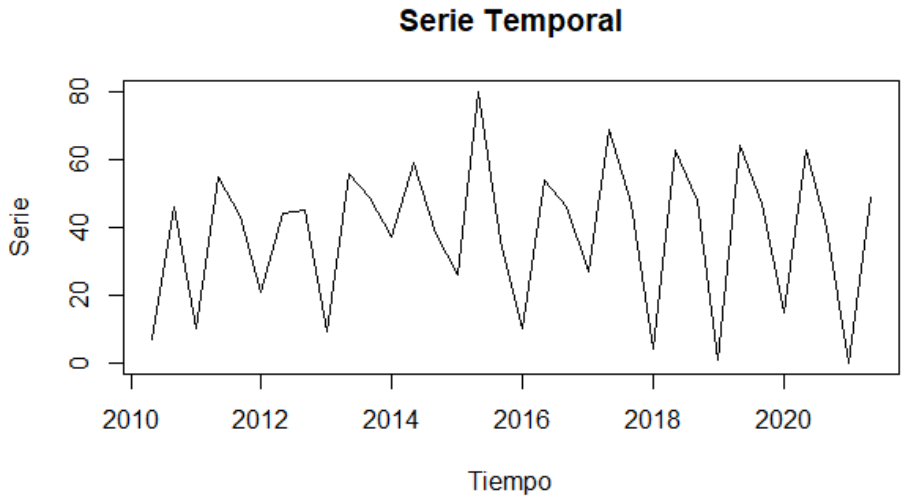
5.1. Análisis de los patrones de la serie de tiempo

A continuación, se presentan los resultados obtenidos para ambos cursos tomados como muestra, además en el anexo 3 y anexo 4 se incluye el script utilizado en R.

5.1.1. Curso de algoritmos y programación

Para el siguiente análisis se tomaron como datos las matrículas en el periodo 2010-I al periodo 2021-I como se muestra en la tabla 18 del anexo 1. Así mismo se puede apreciar en la figura 14, la serie temporal del curso de algoritmos y programación, en la cual se observa que es aditiva.

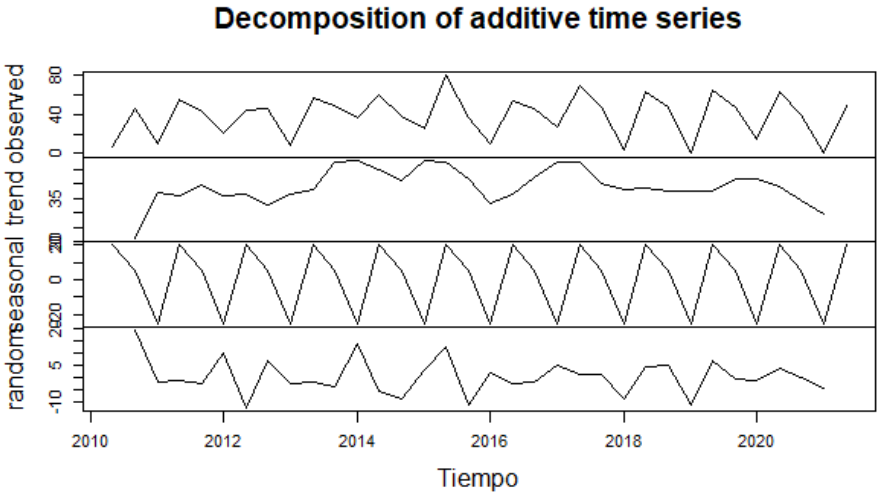
Figura 14 *Serie temporal de matrículas del curso de algoritmos y programación*



Fuente: elaboración propia

Además, aparenta presentar estacionalidad, pero no tendencia; para verificar, se procedió a la descomposición aditiva, la misma que se muestra en la figura 15 en la cual se verifica que existe estacionalidad y tendencia lineal.

Figura 15 *Descomposición aditiva de la serie de cursos de algoritmos y programación*



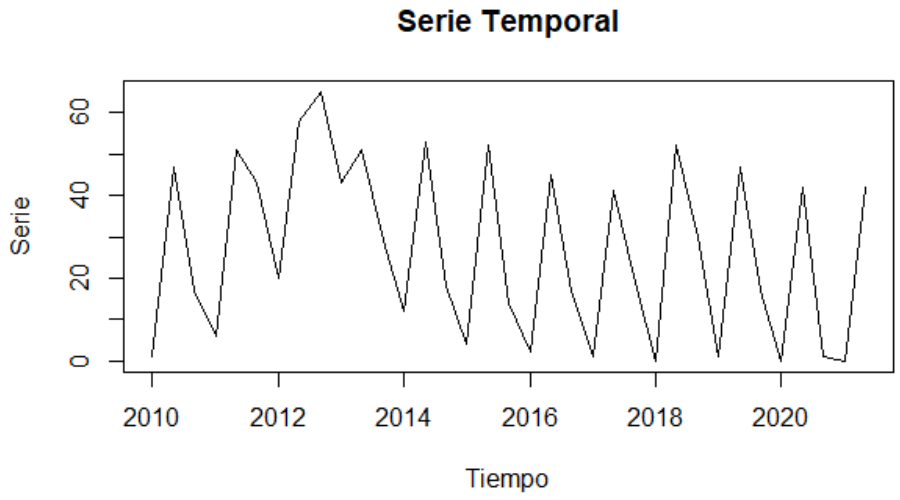
Fuente: Elaboración propia

Nota. Según se verifica en la figura, la existencia de estacionalidad (tercera fila) y de tendencia lineal (segunda fila).

5.1.2. Curso de Calculo III

De la misma manera, para el análisis se tomaron como datos la cantidad de alumnos matriculados en el periodo 2010-0 al periodo 2021-I como se muestra en el cuadro 21 del anexo 2. Así mismo se puede apreciar en la figura 16 la serie temporal del curso de cálculo III, donde se observa que es aditiva.

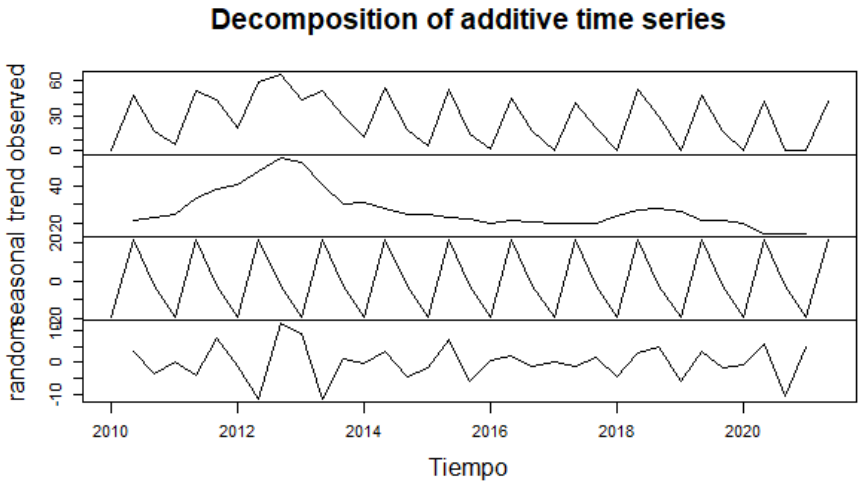
Figura 16 *Serie temporal de matrículas del curso de cálculo III.*



Fuente: elaboración propia

Además, aparenta presentar estacionalidad, pero no tendencia; para verificar, se procedió a la descomposición aditiva, la misma que se muestra en la figura 17, en la cual se verifica que existe estacionalidad y tendencia lineal.

Figura 17 *Descomposición aditiva de la serie de cursos de cálculo III.*



Fuente: elaboración propia

Para ambos cursos se analizarán la descomposición aditiva en la sección 5.2 aplicando los tres modelos.

5.2. Aplicación de modelos

A continuación, se presentan los resultados obtenidos al aplicar los modelos de pronóstico para ambos cursos tomados como muestra los datos de entrenamiento como la serie completa paras cinco y veintidós periodos respectivamente, además en el anexo 3 y anexo 4 se incluye el script utilizado en R.

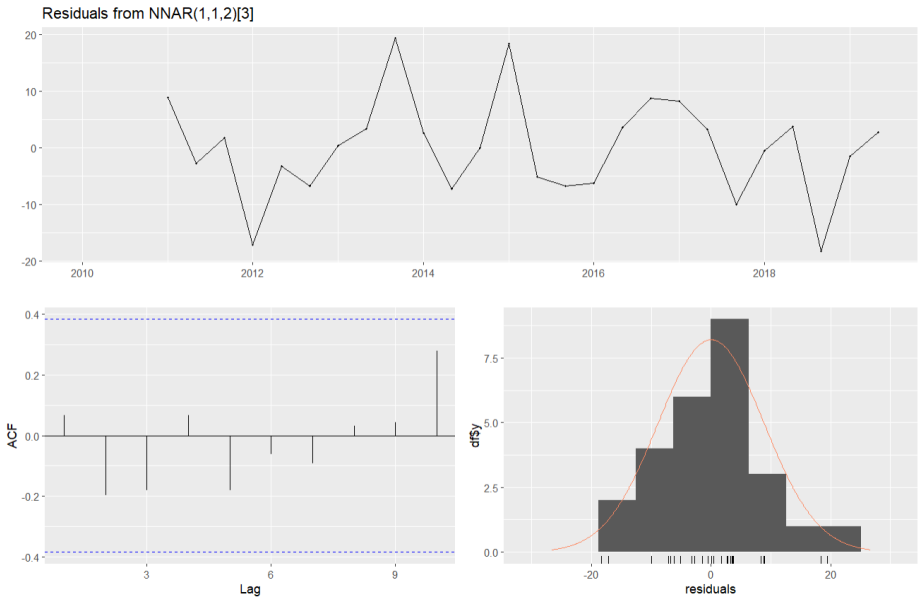
5.2.1. Curso de Algoritmos y programación

A continuación, se evaluará las matrículas en el curso de algoritmos y programación como una serie según el ANEXO 1, las cuales se evaluarán las series de entrenamiento y de control. Se evaluará el pronóstico por los tres métodos de la serie de entrenamiento para cinco periodos (2020-0 al 2021-1), tal que se pueda realizar el control con los datos reales. Como segundo objetivo, se realizará el cálculo del RMSE (error de pronóstico), para cada método y con ello llegará al mejor método que se acerque a los datos reales, es decir, el que tenga mejor calidad de pronóstico. Asimismo, se realizará el pronóstico para veintidós periodos, del 2021-2 al 2028-2 como evaluación de pronóstico de largo plazo en cada método.

5.2.1.1. Red Neuronal

Para este método se aplica la función nnetar (Neural Network Time Series Forecasts) de la librería de pronósticos de R en la cual se creó un modelo de red neuronal de retroalimentación con una sola capa oculta y entradas retrasadas para series de tiempo univariadas, posteriormente se analizaron los residuales (véase en figura 18).

Figura 18 *Análisis residual de la rede neuronal NNAR (1,1,2) [3]*



Fuente: elaboración propia

Como se puede apreciar, los residuales son cercanos a cero, además están dentro de los intervalos de confianza de ACF y seguirían aparentemente una distribución normal. Sin embargo, para tener una certeza de lo anteriormente mencionado se hizo una prueba de normalidad y el test de Ljung-Box a los residuos.

Los resultados de la prueba de normalidad con el test de Shapiro - Wilks se muestran a continuación.

H0: Los residuos siguen una distribución normal

H1: Los residuos no siguen una distribución normal

α: 0,05

W = 0.96427, p-value = 0.4824

En este caso, el valor p es mayor que 0.05 por lo tanto se acepta H0, es decir los residuales siguen una distribución normal.

Con respecto al test de Ljung-Box se obtuvo lo siguiente:

H0: Los residuales se distribuyen de forma independiente

H1: Los residuales no se distribuyen de forma independiente

α: 0.05

Q* = 3.7773, df = 6, p-value = 0.7068

Como se puede apreciar p-value es mayor que 0.05 por lo tanto se acepta H0, es decir, los residuales se distribuyen independientemente. Una vez verificado el cumplimiento de los requisitos necesarios en los residuales, se procederá a calcular el pronóstico para cinco periodos y con ello la validación de la calidad con el indicador RMSE.

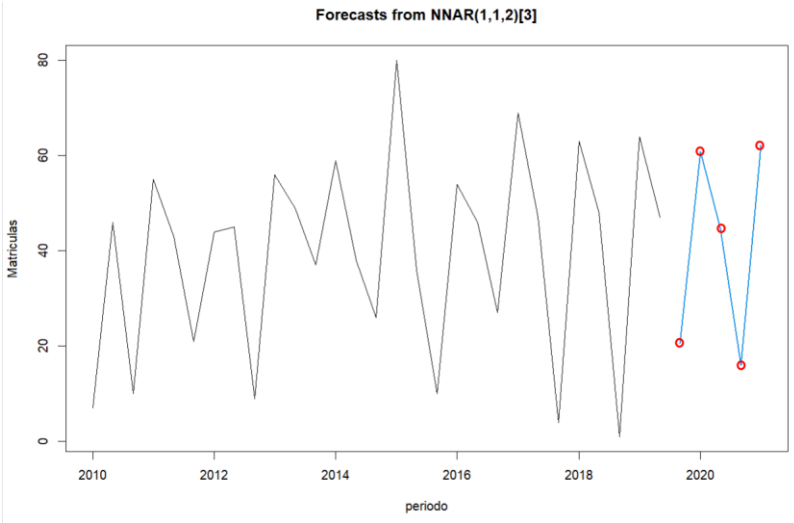
5.2.1.2. Pronósticos por Redes neuronales

A continuación, se evaluarán lo pronostico a cinco periodos (2020-0 al 2021-1) como también a veintidós periodos (2021-2 al 2028-2).

A. Periodo 2020-0 al 2021-1

De acuerdo a lo observado en la figura 19 los pronósticos para los cinco periodos 2020-0 al 2021-1, existe la estacionalidad similar a la serie de entrenamiento como también mantiene la frecuencia trimestral. Asimismo, se realizó el cálculo de la calidad de pronóstico a través RMSE cuyo valor fue 8.697431.

Figura 19 *Pronostico para cinco periodos del curso Algoritmos y programación por el método redes neuronales*



Fuente: elaboración propia

Cuadro 3 *Tabla de pronóstico por el modelo de redes neuronales para cinco periodos del curso Algoritmos y programación según ANEXO 6.*

PERIODO	POINT_FORESCAST	MATRICULAS
2020-0	20.55220	20
2020-1	60.90157	60
2020-2	44.63939	44
2021-0	15.94115	15
2021-1	62.19591	62

Según se observa, en el cuadro 3, los valores de las matrículas son dispersas no repetitivas.

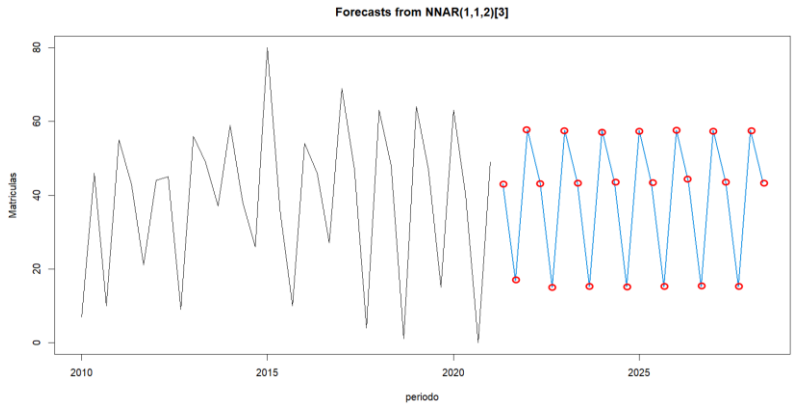
B. Periodo 2021-1 al 2028-2

A continuación, se evalúa la serie de matrículas de la tabla 17 del anexo 1 para el cálculo del pronóstico, aplicando las mismas pruebas de validación de los residuales están dentro del intervalo de confianza ACF y los p-value para el test de Shapiro - Wilks se obtuvo 0.3736, para de Ljung-Box 0.864 ambos mayores a 0.05. Es decir, los residuales se distribuyen independientemente y siguen una distribución normal y se obtuvo lo siguientes resultados en su pronóstico. Asimismo, se obtuvo el RMSE el valor de 8.293674.

De acuerdo a lo observado en la figura 20 se observa que el modelo pronostica valores con una notable estacionalidad, esto podría dase debido a que la serie de aprendizaje también tiene una estacionalidad clara como también no posee tendencia fuerte y poca variación aleatoria

.

Figura 20 *Pronostico para veintidós periodos del curso Algoritmos y programación por el método redes neuronales*



Fuente: elaboración propia

Cuadro 4 *Tabla de pronóstico por el modelo redes neuronales para veintidós periodos del curso algoritmos y programación según ANEXO 6.*

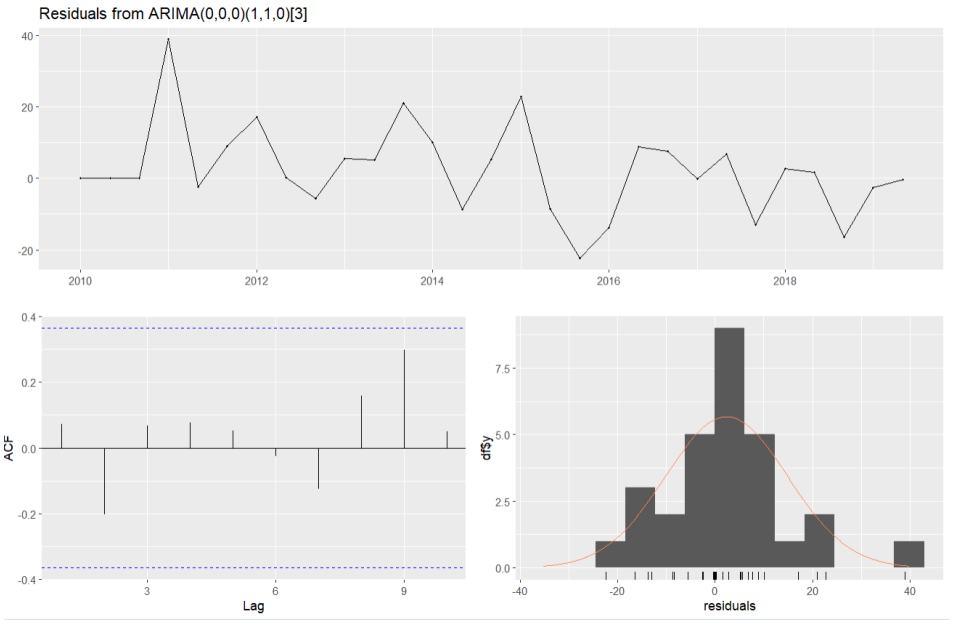
PERIODO	POINT_FORESCAST	MATRICULAS
2021-2	43.05464	43
2022-0	16.80610	16
2022-1	57.66019	57
2022-2	43.39104	43
2023-0	15.08055	15
2023-1	57.65418	57
2023-2	43.41832	43
2024-0	15.19901	15
2024-1	57.64521	57
2024-2	43.42137	43
2025-0	15.19039	15
2025-1	57.64628	57
2025-2	43.42171	43
2026-0	15.19101	15
2026-1	57.64618	57
2026-2	43.42175	43
2027-0	15.64619	15
2027-1	57.64619	57
2027-2	43.42176	43
2028-0	15.19097	15
2028-1	57.64619	57
2028-2	43.42176	43

Así pues, se aprecia en el cuadro 4 el ciclo repetitivo para los valores periódicos a partir del ciclo 2022-1.

5.2.1.3. *Método ARIMA*

Para este modelo se ha utilizado la función auto.arima de R debido a que automáticamente considera el ajuste necesario a una serie con estacionalidad y tendencia, la siguiente figura muestra los resultados del análisis de residuales.

Figura 21 *Análisis residual por el método ARIMA*



Fuente: elaboración propia

Se observa que los residuos también son cercanos a cero, además están dentro del intervalo de confianza en ACF y aparentemente tendrían una distribución normal. Para confirmar la normalidad se aplicó la prueba de Shapiro-Wilks, cuyos resultados se muestran a continuación.

H0: Los residuos siguen una distribución normal

H1: Los residuos no siguen una distribución normal

α: 0,05

W = 0.95472, p-value = 0.2422

Así pues, el valor p es mayor que 0.05 por lo tanto se acepta H0, es decir los residuales siguen una distribución normal. Posteriormente se verificó la independencia de los residuos mediante la prueba de Ljung-Box, cuyos resultados se muestran a continuación.

H0: Los residuales se distribuyen de forma independiente

H1: Los residuales no se distribuyen de forma independiente

α: 0.05

Q* = 2.0192, df = 5, p-value = 0.8465

Se observa que p-value es mayor que 0.05 por lo tanto se acepta H0, es decir, los residuales se distribuyen independientemente. Es decir, este modelo es apto para este los pronósticos de la serie que se evaluaran en la sección siguiente.

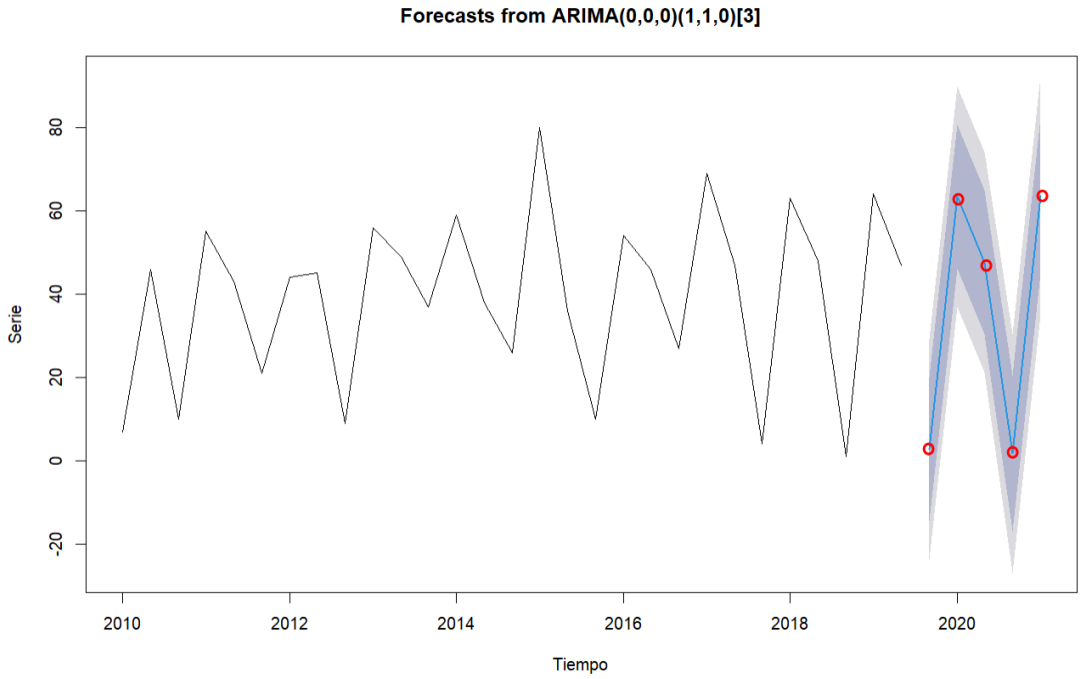
5.2.1.4. Pronósticos por el método ARIMA

A continuación, se realizó la evaluación de pronóstico para cinco periodos del (2020-0 al 2021-1) como también para veintidós periodos (2021-2 al 2028-2).

A. Periodo 2020-0 al 2021-1

Se observa en las regiones sombreadas los limites superiores e inferiores (sombreado oscuro 80% de probabilidad y sombreado claro un 95% de probabilidad de encontrar un dato real, de acuerdo a los datos de pronóstico en el anexo 6). Podemos ver que a comparación del método red neuronal los valores más bajos se ajustan más o tienen mayor peso de los últimos valores de la serie. Así pues, se realizó el cálculo de la calidad de pronóstico a través RMSE cuyo valor fue 12.52304. Este indicador es calculado en base a la función accuracy (pronostico, test), la cual indica la raíz cuadrada del error cuadrático medio entre los datos pronosticados y reales, la cual se encuentra plasmada en el SCRIPT del anexo 3.

Figura 22 *Pronostico de cinco periodos del curso Algoritmos y programación por el método ARIMA*



Fuente: elaboración propia

Cuadro 5 *Tabla de pronóstico por el modelo ARIMA para cinco periodos del curso Algoritmos y programación, según ANEXO 6*

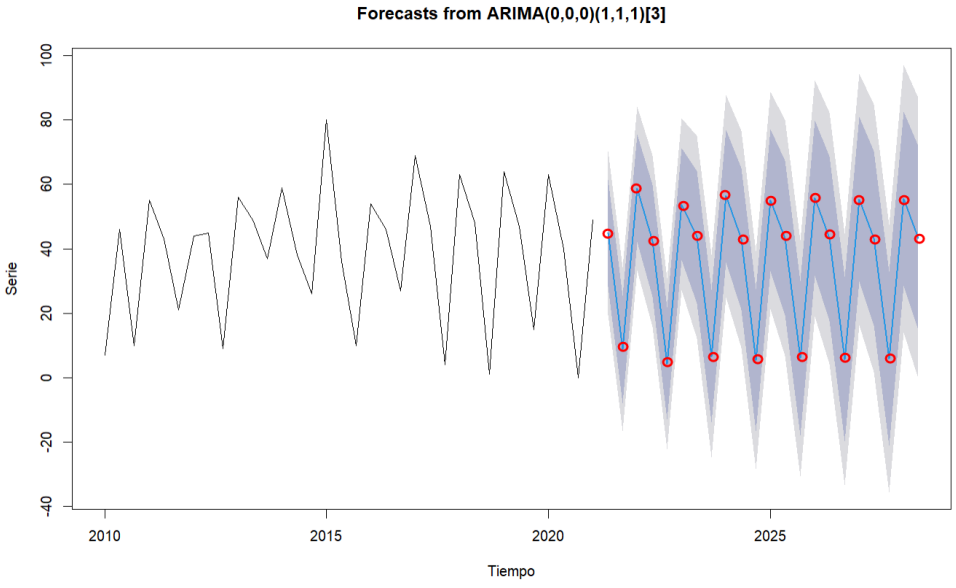
PERIODO	POINT_FORESCAST	MATRICULAS
2020-0	2.753632	2
2020-1	63.415456	63
2020-2	47.584544	47
2021-0	1.728557	1
2021-1	63.757148	64

Como se puede apreciar en el cuadro 5, los valores pronosticados para los cinco periodos no presentan valores repetidos presentando aleatoriedad.

B. Periodo 2021-2 al 2028-2

Así también, para la evaluación de pronóstico se aplicaron las mismas pruebas de validación de los residuales que están dentro del intervalo de confianza ACF y los p-value para el test de Shapiro - Wilks se obtuvo 0.2192, para de Ljung-Box 0.6858 ambos mayores a 0.05. Es decir, los residuales se distribuyen independientemente y siguen una distribución normal el valor RMSE fue 11.9793. Asimismo, se puede observar en la figura 23 que presenta estacionalidad y frecuencia notable a la serie real, en las regiones sombreadas los limites superiores e inferiores (sombreado oscuro 80% de probabilidad y sombreado claro un 95% de probabilidad de encontrar un dato real, de acuerdo a los datos de pronóstico en el anexo 6). Además, se calculó la calidad de pronóstico a través de la función sqrt(mean(residuos^2).

Figura 23 *Pronostico de veintidós periodos del curso Algoritmos y programación por el método ARIMA*



Fuente: elaboración propia

Cuadro 6 Tabla de pronóstico por el modelo ARIMA para veintidós periodos del curso Algoritmos y programación según ANEXO 6.

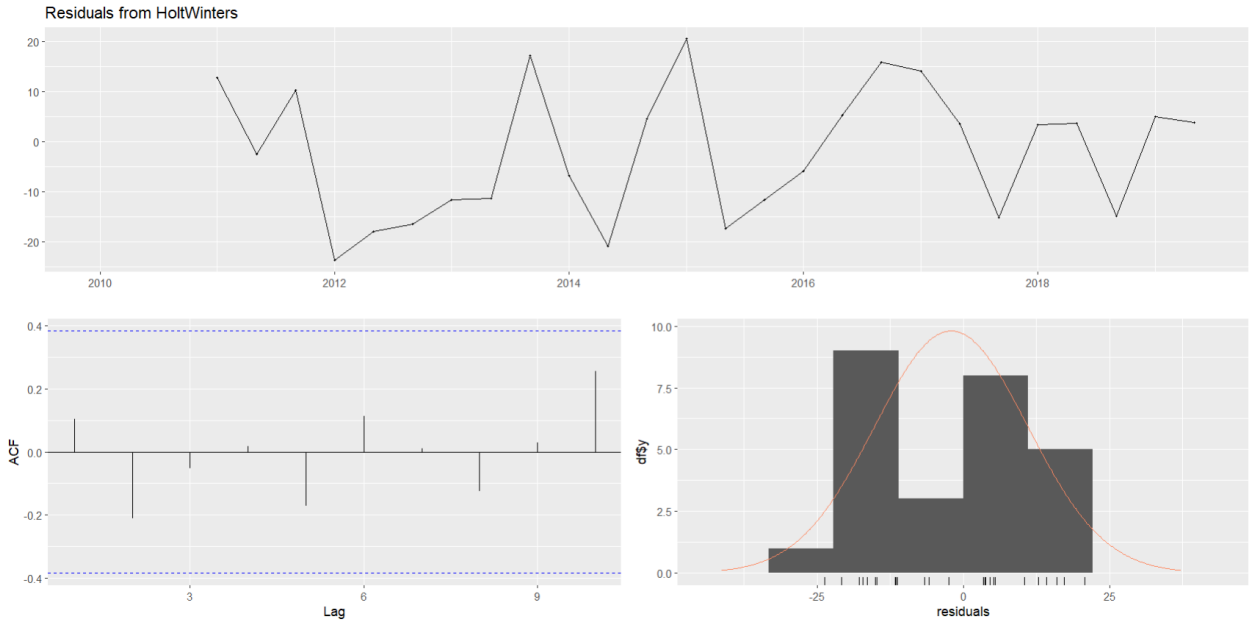
PERIODO	POINT_FORESCAST	MATRICULAS
2021-2	44.94718	44
2022-0	8.756254	8
2022-1	58.831512	58
2022-2	42.438694	42
2023-0	4.316364	4
2023-1	53.846408	53
2023-2	43.710631	43
2024-0	6.567626	6
2024-1	56.374123	56
2024-2	43.065691	43
2025-0	5.426115	5
2025-1	55.092436	55
2025-2	43.39271	43
2026-0	6.004922	6
2026-1	55.74232	55
2026-2	43.226894	43
2027-0	5.711436	5
2027-1	55.412794	55
2027-2	43.310972	43
2028-0	5.860249	5
2028-1	55.579882	55
2028-2	43.26834	43

Como se puede observar en el cuadro 6, presenta valores periódicos a partir del 2026-1, la variabilidad se mantiene hasta los primeros 5 trimestres, luego la serie presenta mayor estacionalidad.

5.2.1.5. *Método Holt-Winters*

Para la prueba de la misma serie de prueba con este modelo se utilizó la función HW de R para la generación del modelo, la siguiente figura muestra el análisis residual, en donde se aprecia que similar a los casos anteriores, los residuos son cercanos a cero, están dentro de los intervalos de confianza del diagrama ACF y gráficamente mostrarían una distribución normal.

Figura 24 *Análisis residual por el método Holt-Winters*



Fuente: elaboración propia

Para confirmar la normalidad se aplicó la prueba de Shapiro-Wilks, cuyos resultados se muestran a continuación.

H0: Los residuos siguen una distribución normal

H1: Los residuos no siguen una distribución normal

α: 0,05

W = 0.94596, p-value = 0.1863

En este caso, el valor p es mayor que 0.05 por lo tanto se acepta H0, es decir los residuales siguen una distribución normal.

Con respecto al test de Ljung-Box se obtuvo lo siguiente:

H0: Los residuales se distribuyen de forma independiente

H1: Los residuales no se distribuyen de forma independiente

α: 0.05

Q* = 3.282, df = 6, p-value = 0.7727

Como se puede apreciar p-value es mayor a 0.05, por lo tanto, se podría aceptar H0, es decir, los residuales se distribuyen independientemente. Este modelo cumple con los requisitos, se podría aplicar para la serie.

5.2.1.6. Pronósticos por el método HOLT-WINTERS

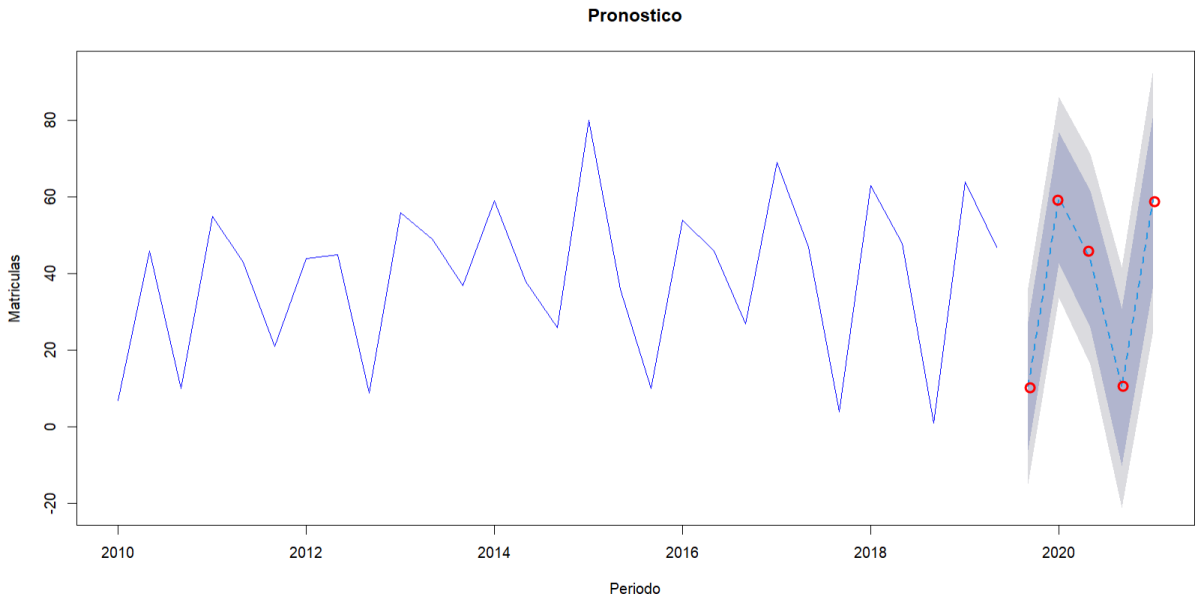
A continuación, se evalúa los pronósticos para cinco periodos del (2020-0 al 2021-1) como también para veintidós periodos (2021-2 al 2028-2).

A. Periodo 2020-0 al 2021-1

Para la evaluación de pronóstico por este modelo de la misma forma se toman los datos de entrenamiento presentado en el anexo 1, los pronósticos a calcular serán para cinco periodos del 2020-0 al 2021-1 obteniendo los siguientes resultados según la figura siguiente.

Tal como se puede apreciar en la figura 25 los valores pronosticados de la serie de entrenamiento guardan relación en cuanto estacionalidad y periodo con la serie de entrenamiento también presenta los limites en la región sombreada, la más cercada con un 80% de probabilidad y la más lejana con 95% de probabilidad de encontrar un próximo valor real, de la misma manera se aplicó el RMSE, para hallar la calidad cuyo valor fue de 12.995555.

Figura 25 *Pronostico de cinco periodos por el método Holt-Winters de matrículas del curso de algoritmos y programación.*



Fuente: elaboración propia

Cuadro 7 *Tabla de pronóstico para cinco periodos del curso de algoritmos y programación por el método HOLT-WINTERS de acuerdo al ANEXO 6.*

PERIODO	POINT_FORESCAST	MATRICULAS
2020-0	10.74290	10
2020-1	59.89706	59
2020-2	43.82668	43
2021-0	10.31806	10
2021-1	59.47223	59

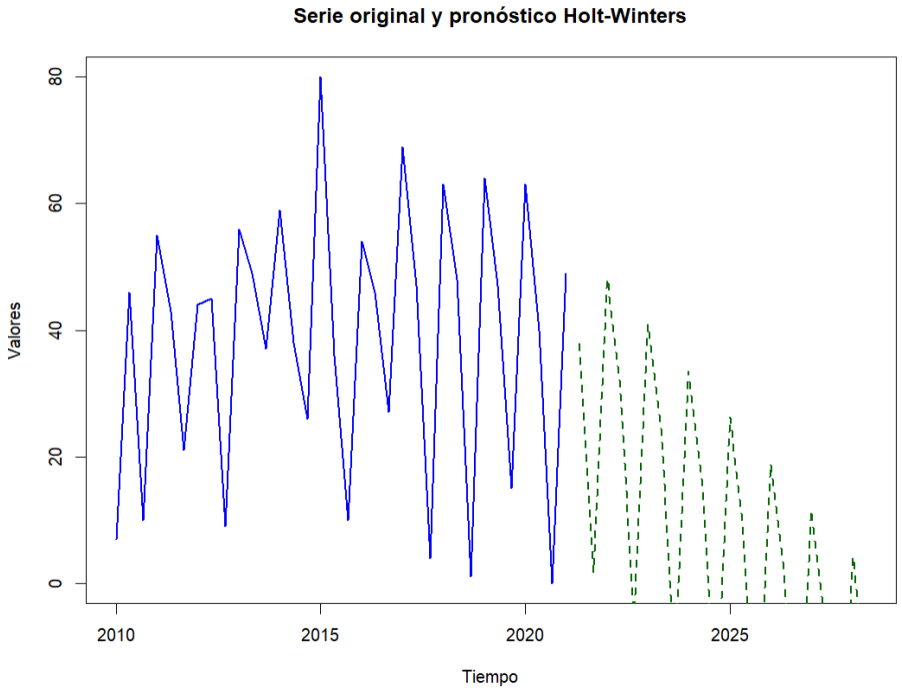
De acuerdo a lo que se puede observar en el cuadro 7, los valores pronosticados guardan similitud a los otros métodos a pesar de tener un RMSE alto.

B. Periodo 2021-2 al 2028-2

Por otro lado, para realizar el análisis de pronóstico para veintidós periodos (h=22), también se efectuó las mismas pruebas de validación de los residuales que están dentro del intervalo de confianza ACF y los p-value para el test de Shapiro - Wilks se obtuvo 0.3652, para de Ljung-Box 0.6527 ambos mayores a 0.05. Es decir, los residuales se distribuyen independientemente y siguen una distribución normal cuyos datos se muestran a continuación.

Respecto a la evaluación de pronóstico, se hallaron valores negativos. Al evaluar restricciones, se tendrían que ajustar manualmente a cero, no recomendable para la evaluación estadística como también podría distorsionar los intervalos. En este sentido, se procedió también a analizar un truncamiento logarítmico para lograr restringir los pronósticos negativos en los pronósticos con la aplicación log1p() + Holt-Winters aditivo, en la que el pronóstico obteniendo RMSE fue muy pobre para la cual en la presente investigación se optó por continuar trabajando con los pronósticos negativos y posterior evaluación tal como muestra la figura 28 la cual presenta mejor calidad. Asimismo, se calculó el RMSE cuyo valor fue de 12.29484, la cual será comparada con los otros métodos advirtiendo que evalúa valores negativos.

Figura 26 *Pronostico de veintidós periodos por el método Holt- winters del curso Algoritmos y programación.*



Fuente: elaboración propia

Cuadro 8 *Tabla de pronóstico por el método Holt-Winters para veintidós periodos del curso algoritmos y programación según el ANEXO 6.*

PERIODO	POINT_FORESCAST	MATRICULAS
2021-2	37.827639	37
2022-0	1.704127	1
2022-1	48.211559	48
2022-2	30.492766	30
2023-0	-5.630746	0
2023-1	40.876685	40
2023-2	23.157892	23
2024-0	-12.965619	0
2024-1	33.541812	33

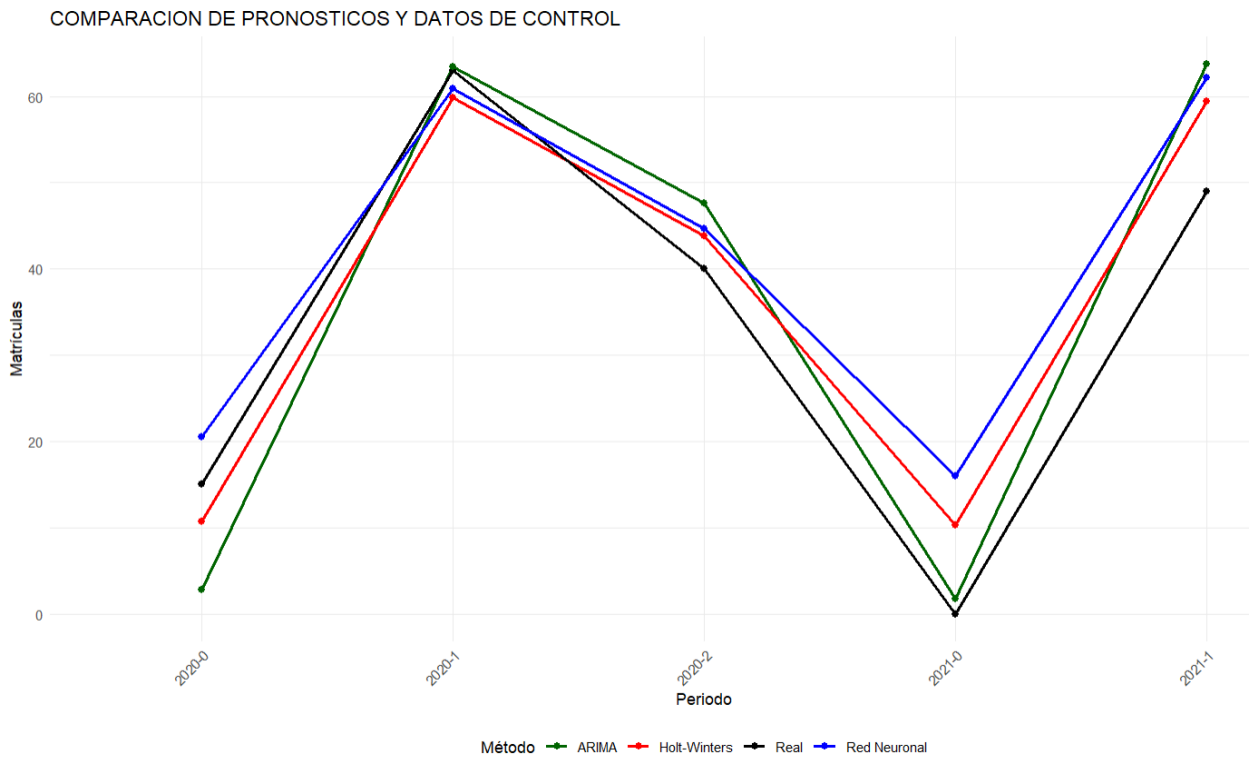
2024-2	15.823019	15
2025-0	-20.300493	0
2025-1	26.206939	26
2025-2	8.488145	8
2026-0	-27.635366	0
2026-1	18.872065	18
2026-2	1.153272	1
2027-0	-34.97024	0
2027-1	11.537192	11
2027-2	-6.181601	0
2028-0	-42.305113	0
2028-1	4.202318	4
2028-2	-13.516475	0

En el cuadro 8 se visualiza los pronósticos negativos estimados de las matrículas, la cual requiere ajuste manual para considerarlos con valores igual a cero.

5.2.1.7 Selección del modelo

Para la selección del modelo se utilizó como criterio el indicador RMSE. Así pues, se presentarán los resultados para los pronósticos de los periodos (2020-0 al 2021-1), las cuales esta descritos y comparados por cada método en la figura 27, como también la comparación de los resultados de calidad en el cuadro 9. Posterior, se describen los resultados al pronosticar los periodos (2021-2 al 2028-2) en la figura 28, en la que se comparan los valores pronosticados por los tres métodos, como también la comparación del indicador de calidad en el cuadro 9, para la comparación y elección del modelo.

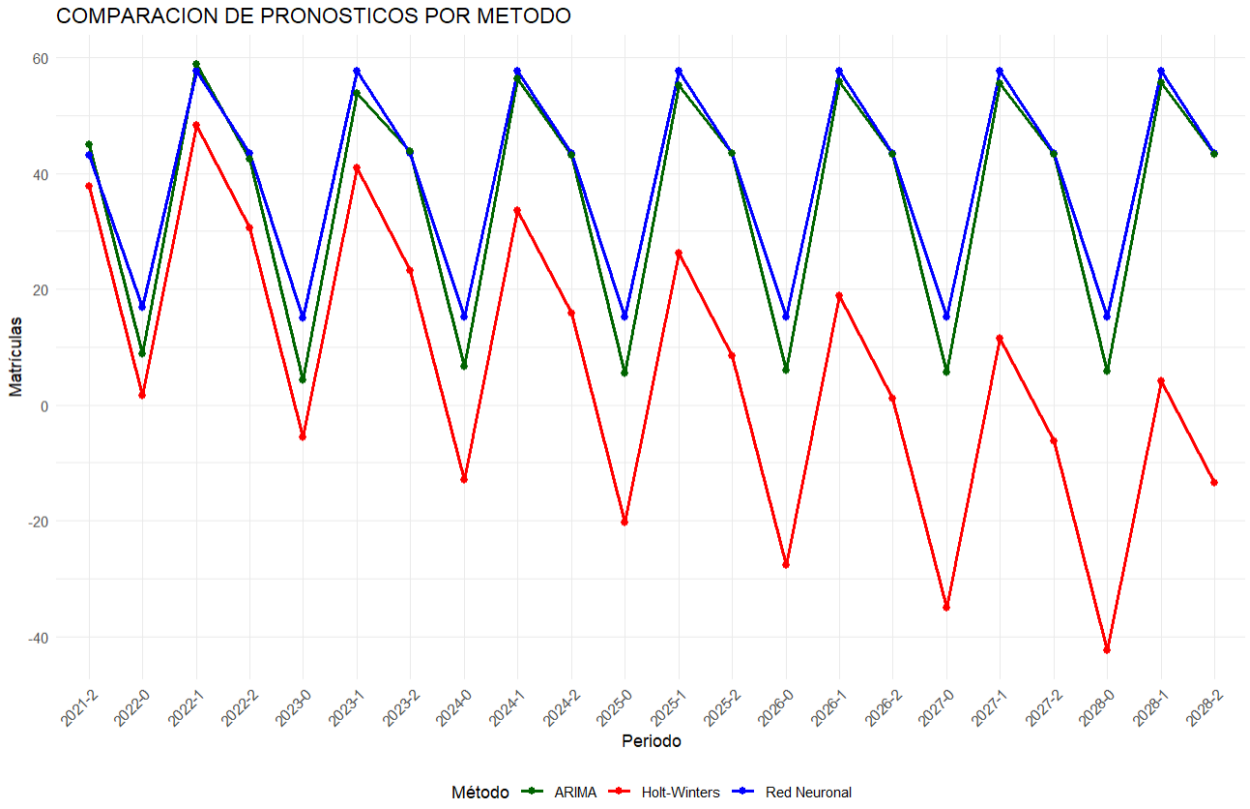
Figura 27 *Pronostico de los periodos (2020-0 al 2021-1) por los tres métodos para el curso algoritmos y programación.*



Fuente: elaboración propia

Según se puede observar en la figura 27, que el método ARIMA y de Redes Neuronales se encuentra más cercano a la serie de control, esto se validará con el estadístico de calidad RMSE.

Figura 28 *Pronostico de los periodos (2021-1 al 2028-2) por los tres métodos para el curso algoritmos y programación.*



Fuente: elaboración propia

Se puede visualizar que los métodos de Redes Neuronales y ARIMA son las que tienen mayor similitud, como también presentan valores mayores a cero acercándose mejor a las características de la serie. Por otro lado, el método Holt-Winters, tiene los limites más alejados incluso menores a cero, datos que no comparte similitud con la serie. Esto pues, se verá reflejado

en el valor RMSE como indica el cuadro 9 en la que se compararon las calidades en cada aplicación de método y la de Holt- Winters fue la de más baja calidad.

Cuadro 9 *Tabla de resultados RMSE de los tres métodos para el curso algoritmos y programación.*

Método/RMSE	**RMSE (h=5)**	**RMSE(h=22)**
Redes Neuronales	8.697431	8.293674
Arima	12.52304	11.9793
Holt-Winters	12.995555	12.29484

Se concluye que, la mejor calidad para los pronósticos del periodo (2020-1 al 2021-1) y del periodo (2021-2 al 2028-2) es el del método Redes Neuronales, cuyo valor RMSE fue el menor que los otros dos métodos.

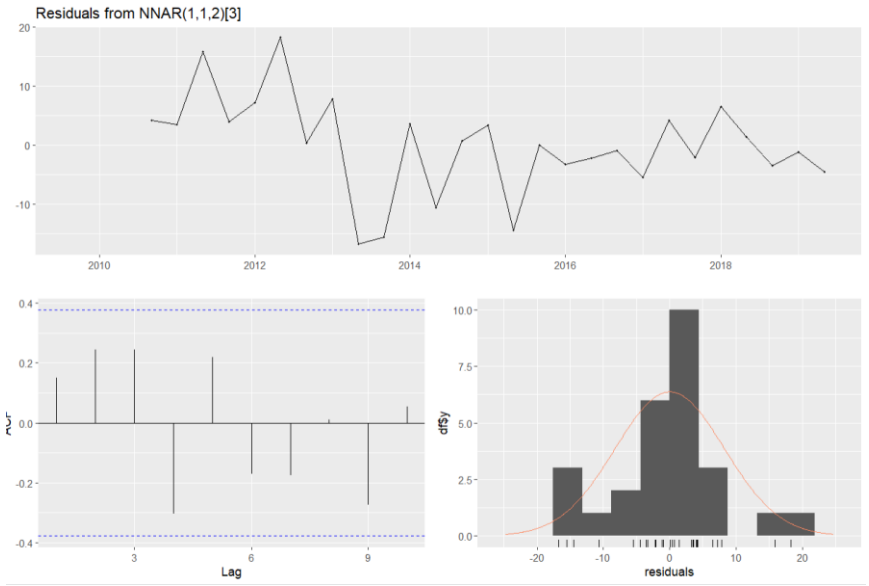
5.2.2. Curso de Calculo III

A continuación, se evaluará el pronóstico aplicando los tres métodos para la serie de datos de entrenamiento plasmados en el ANEXO 2. El pronóstico se realiza para cinco periodos (2020-0 al 2021-1), tal que se pueda realizar el control con los datos de control. Posteriormente, se realizará el pronóstico para veintidós periodos (2021-2 al 2028-2) en la que también se evaluará la calidad de pronóstico de cada método.

5.2.2.1. *Red Neuronal*

En la siguiente prueba se realizó los pronósticos de la serie de prueba (test) para cinco periodos del anexo 2. Mediante la función nnetar (Neural Network Time Series Forecasts) de la librería de pronósticos de R, se creó un modelo de red neuronal de retroalimentación con una sola capa oculta y entradas retrasadas para series de tiempo univariadas, posteriormente se analizaron los residuales, las que se detallan a continuación.

Figura 29 *Análisis residual por Red Neuronal*



Fuente: elaboración propia

Como se puede apreciar, los residuales son cercanos a cero, además están dentro de los intervalos de confianza de ACF y seguirían una distribución normal. Sin embargo, para una confirmación de lo anteriormente mencionado se hizo una prueba de normalidad y el test de Ljung-Box a los residuos.

Los resultados de la prueba de normalidad con el test de Shapiro - Wilks se muestran a continuación.

H0: Los residuos siguen una distribución normal

H1: Los residuos no siguen una distribución normal

α: 0,05

W = 0.95432, p-value = 0.2723

En este caso, el valor p es mayor que 0.05 por lo tanto se acepta H0, es decir los residuales siguen una distribución normal.

Con respecto al test de Ljung-Box se obtuvo lo siguiente:

H0: Los residuales se distribuyen de forma independiente

H1: Los residuales no se distribuyen de forma independiente

α: 0.05

Q* = 10.44, df = 6, p-value = 0.1073

Como se evalúa al apreciar p-value es mayor que 0.05 por lo tanto se acepta H0, es decir, los residuales se distribuyen independientemente. Una vez verificado el cumplimiento de los requisitos necesarios se procederá a realizar el cálculo del RMSE para cada uno de los pronósticos.

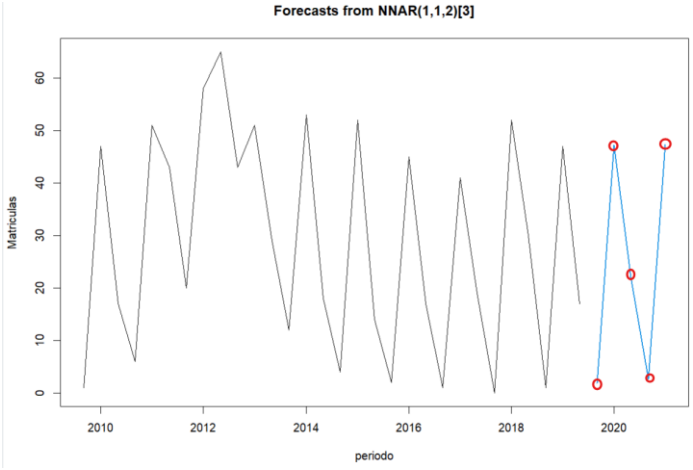
5.2.2.2. Pronósticos por Redes neuronales

A continuación, se evaluarán lo pronostico a cinco periodos (2020-0 al 2021-1) como también a veintidós periodos (2021-2 al 2028-2).

A. Periodo 2020-0 al 2021-1

De acuerdo a lo observado en la figura 30, se evidencia el modelo aplicado NNAR(1,1,2)[3], el cual tiene como numero de retardos no estacionales AR(1) igual a uno, numero de retardo estacional SAR (1) también uno, dos neuronas en la capa oculta y una estacionalidad igual a tres, según se analizó la serie. En seguida, se procedió con el cálculo del RMSE cuyo valor fue de 8.090142.

Figura 30 *Pronostico (2020-0 al 2021-1) del curso calculo III por el método redes neuronales*



Fuente: elaboración propia

Cuadro 10 *Tabla de pronóstico por el modelo de redes neuronales para cinco periodos del curso calculo III de acuerdo al ANEXO 5*

PERIODO	POINT_FORESCAST	MATRICULAS
2020-0	1.838517	1
2020-1	47.184502	47
2020-2	22.014653	22
2021-0	2.659155	2
2021-1	47.303200	47

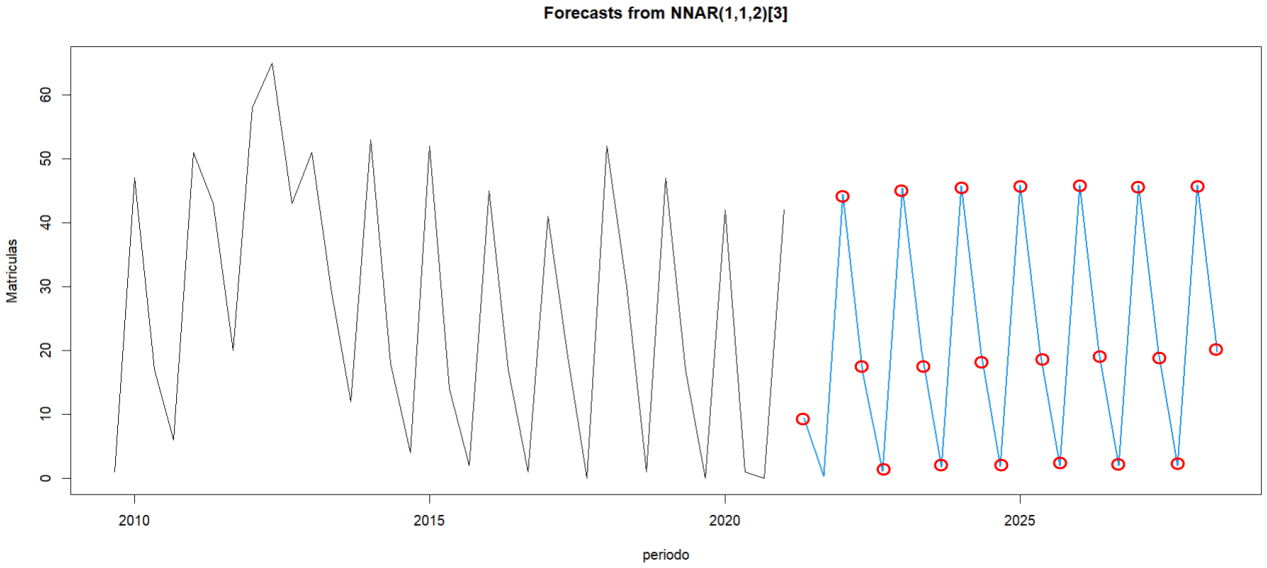
EL cuadro 10 representa los pronósticos presentados de Rstudio en el anexo 5 las cuales presentan estacionalidad.

B. Periodo 2021-2 al 2028-2

Por otro lado, se evalúa la serie veintidós periodos (2021-2 al 2028-2) en la cual se podrán observar un patrón repetitivo, como también la viabilidad del modelo con los indicadores de prueba.

Como se puede apreciar en la figura 31, los pronósticos para h=22 periodos, también siguen el mismo modelo NNAR(1,1,2)[3]. Así también, se aplicó las mismas pruebas de validación de los residuales que están dentro del intervalo de confianza ACF y los p-value para el test de Shapiro - Wilks se obtuvo 0.4824, para de Ljung-Box 0. ambos mayores a 0.05. Es decir, los residuales se distribuyen independientemente y siguen una distribución normal. Asimismo, se calculó el RMSE cuyo valor fue 8.050362.

Figura 31 *Pronostico (2021-1 al 2028-2) del curso calculo III por el método redes neuronales*



Fuente: elaboración propia

Cuadro 11 *Pronóstico (2021-1 al 2028-2) por el modelo redes del curso calculo III según ANEXO 5.*

PERIODO	POINT_FORESCAST	MATRICULAS
2021-2	12.524251	12
2022-0	0.798022	0
2022-1	43.913345	43
2022-2	16.956572	16
2023-0	1.329073	1
2023-1	45.091502	45
2023-2	17.789332	17
2024-0	1.498055	1
2024-1	45.617264	45
2024-2	18.124345	18
2025-0	1.563592	1
2025-1	45.813646	45
2025-2	18.229422	18
2026-0	1.586311	1

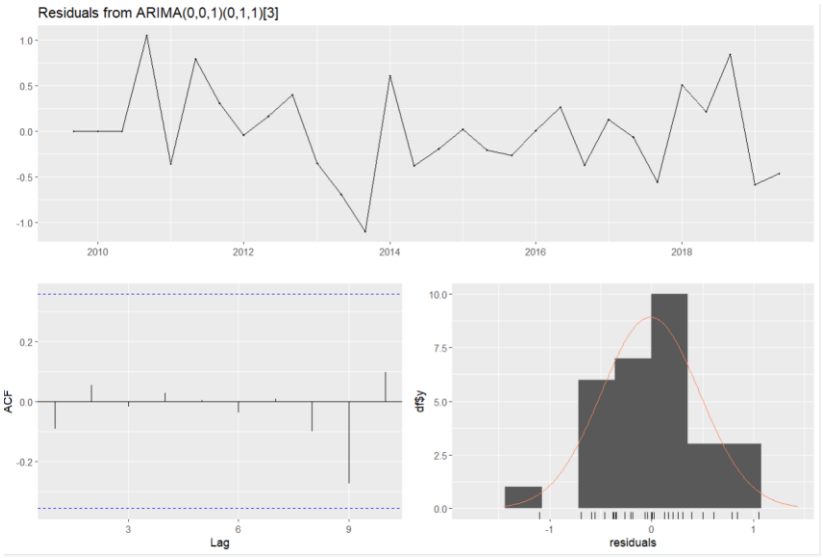
2026-1	45.881626	45
2026-2	18.270051	18
2027-0	1.594783	1
2027-1	45.904569	45
2027-2	18.281229	18
2028-0	1.597435	1
2028-1	45.912227	45
2028-2	18.286090	18

Como se evidencia en el cuadro 11 existen valores repetitivos a partir del periodo 2024-0, característica también de la serie del otro curso.

5.2.2.3. *Método ARIMA*

La aplicación se realizó para el pronóstico de la serie de entrenamiento del anexo 2 para cinco periodos, la cual será comparada con los otros métodos posteriormente. Para este caso se ha utilizado la función auto.arima de R ya que automáticamente considera el ajuste necesario a una serie con estacionalidad y tendencia, la figura 32 se muestra los resultados del análisis de residuales.

Figura 32 *Análisis residual por el método ARIMA*



Fuente: elaboración propia

Se observa que los residuos también son cercanos a cero, además están dentro del intervalo de confianza en ACF y al parecer tendrían una distribución normal. Para confirmar la normalidad se aplicó la prueba de Shapiro-Wilks cuyos resultados se muestran a continuación

H0: Los residuos siguen una distribución normal

H1: Los residuos no siguen una distribución normal

α: 0,05

W = 0.98548, p-value = 0.9451

Así pues, el valor p es mayor que 0.05 por lo tanto se acepta H0, es decir los residuales siguen una distribución normal. Posteriormente se verificó la independencia de los residuos mediante la prueba de Ljung-Box, cuyos resultados se muestran a continuación.

H0: Los residuales se distribuyen de forma independiente

H1: Los residuales no se distribuyen de forma independiente

α: 0.05

Q* = 0.47859, df = 4, p-value = 0.9756

Se observa que p-value es mayor que 0.05 por lo tanto se acepta H0, es decir, los residuales se distribuyen independientemente.

5.2.2.4. Pronósticos por el método ARIMA

A continuación, se evaluarán los pronósticos para el curso de CALCULO III presentado en el ANEXO 5, en esta ocasión por el método ARIMA, en primer análisis para cinco periodos y posterior para veintidós periodos.

A. Periodo 2020-0 al 2021-1

Para el pronóstico h= 5 por el método ARIMA, de acuerdo al Script del anexo 4, se usó la función auto.arima() para hallar el mejor método ARIMA aplicable a la serie con posibilidad que la serie tenga estacionalidad seasonal= TRUE, con ello también explorar la posibilidad de aplicar método SARIMA, el modelo adecuado seleccionado fue ARIMA(0,0,1)(0,1,1)[3], un método ARIMA con estacionalidad, frecuencia igual a tres, es decir SARIMA. Interpretando el modelo a continuación.

ARIMA (0,0,1)(0,1,1)[3]

- No estacionaria (0,0,1)

  P=0, no presenta componente autorregresivo AR(o)

  d=o, no aplica diferenciación no estacional

  q=1, aplica un componente MA, promedio móvil de orden uno

- Estacionaria (0,1,1) [3]

  p=0, sin componentes AR estacional

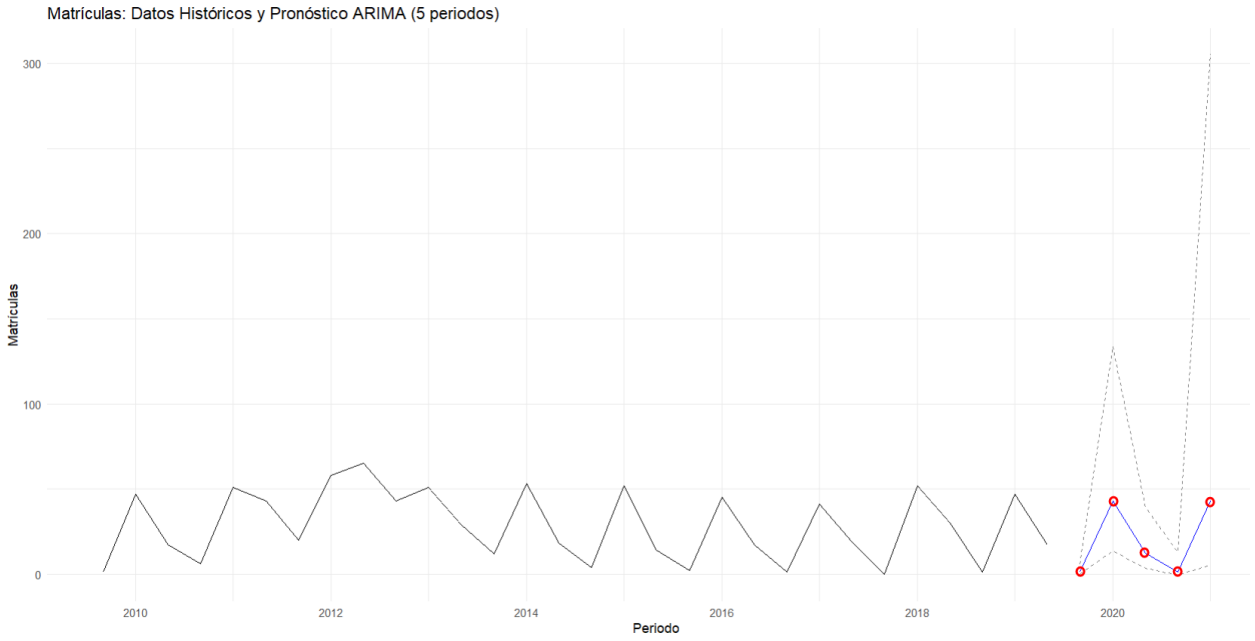
  D=1, se aplica una diferencia estacional, se usa para estabilizar la estacionalidad

  Q=1, hay un componente promedio móvil estacional de orden uno

[3]: la frecuencia estacional es 3, lo indica en la serie(trimestral).

El modelo aplicado finalmente es un SARIMA, puesto a la serie tiene comportamiento estacional (frecuencia 3), la serie no tiene tendencia fuerte, por lo que no requiere ajustes no estacionales. Así también, se el valor del RMSE para cinco periodos fue de 5.138969

Figura 33 *Pronóstico de cinco periodos por el modelo ARIMA para el curso de cálculo III*



Fuente: elaboración propia

Cuadro 12 *Tabla de pronóstico para cinco periodos del curso calculo III por el método ARIMA, según ANEXO 5*

PERIODO	POINT_FORESCAST	MATRICULAS
2020-0	1.407766	1
2020-1	43.041724	43
2020-2	12.244673	12
2021-0	1.204221	1
2021-1	43.041724	43

El cuadro 12 indica los valores de pronóstico los siguientes periodos 2020-0 al 2021-1. Asimismo, para estos valores se puede observar sombreado los limites inferior y superior en la figura 33 de las probabilidades de encontrar un dato real al pronosticar es de 95% de probabilidad marcado por la línea punteada.

B. Periodo 2021-2 al 2028-2

Para el pronóstico de veintidós periodos, de acuerdo al Script del anexo 4, el modelo adecuado seleccionado fue ARIMA (0,0,0) (0,1,0) [3] como se aprecia en la figura 34.

- No estacionaria (0,0,0)

  P=0, no presenta componente autorregresivo AR(o)

  d=o, no aplica diferenciación no estacional

  q=0, no aplica un componente MA

- Estacionaria (0,1,0) [3]

p=0, sin componentes AR estacional

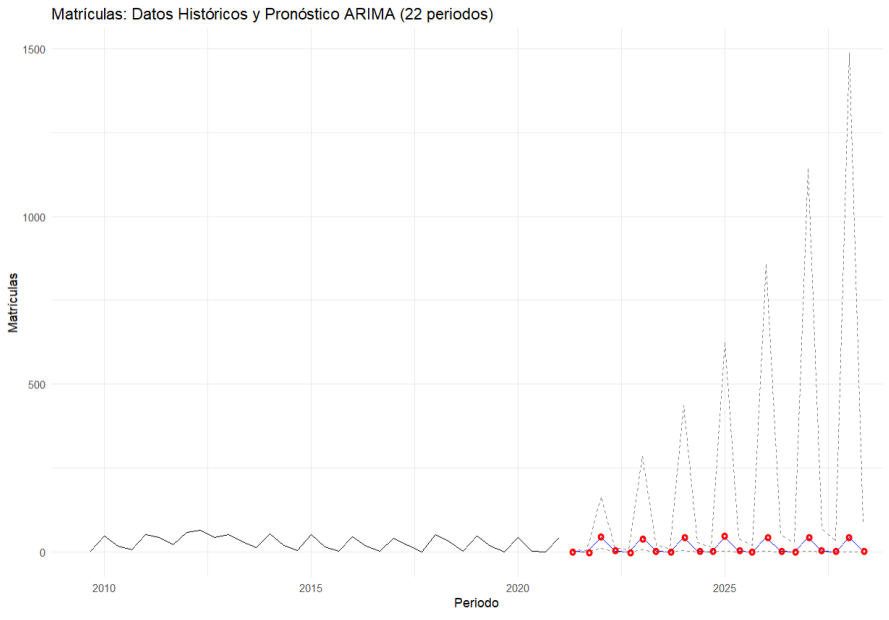
d=1, se aplica una diferencia estacional, se usa para estabilizar la estacionalidad

Q=0, no hay un componente promedio móvil estacional.

[3]: la frecuencia estacional es 3, lo indica en la serie(trimestral).

Aplicando las mismas pruebas de validación de los residuales que están dentro del intervalo de confianza ACF y los p-value para el test de Shapiro - Wilks se obtuvo 0.06734, para de Ljung-Box 0.2073 ambos mayores a 0.05. Es decir, los residuales se distribuyen independientemente y siguen una distribución normal y se obtuvo lo siguientes resultados en su pronóstico. Asimismo, se calculó el RMSE cuyo valor fue 12.29406.

Figura 34 *Pronostico de veintidós periodos por el método ARIMA para el curso de cálculo III.*



Fuente: elaboración propia

Se visualiza que los límites en los pronósticos están muy por encima que los datos de la serie original y poco probable que un dato se encuentre en los limites superiores, estas se encuentran representadas por la línea punteada con 95% de probabilidad de encontrar un dato real al pronosticarlo.

Cuadro 13 *Tabla de pronóstico por el método ARIMA para veintidós periodos del curso calculo III según ANEXO 5*

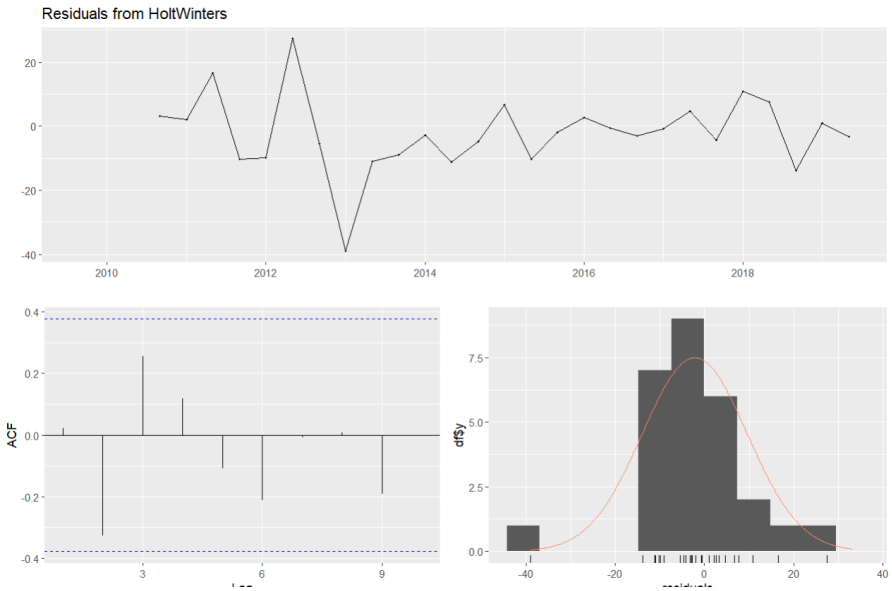
PERIODO	POINT_FORESCAST	MATRICULAS
2021-2	1	1
2022-0	0	0
2022-1	42	42
2022-2	1	1
2023-0	0	0
2023-1	42	42
2023-2	1	1
2024-0	0	0
2024-1	42	42
2024-2	1	1
2025-0	0	0
2025-1	42	42
2025-2	1	1
2026-0	0	0
2026-1	42	42
2026-2	1	1
2027-0	0	0
2027-1	42	42
2027-2	1	1
2028-0	0	0
2028-1	42	42
2028-2	1	1

Se puede visualizer en el cuadro 13 el ciclo notable para los valores pronosticados, notándose una notable estacionalidad.

5.2.2.5. *Método HOLT-WINTERS*

Para la prueba con este modelo se utilizó la función HW de R para la generación del modelo, la figura 35 muestra el análisis residual, en donde se aprecia que similar a los casos anteriores, los residuos son cercanos a cero, están dentro de los intervalos de confianza del diagrama ACF y gráficamente mostrarían una distribución normal

Figura 35 *Análisis residual por el método Holt-Winters*



Fuente: elaboración propia.

Para confirmar la normalidad se aplicó la prueba de Shapiro-Wilks cuyos resultados se muestran a continuación.

H0: Los residuos siguen una distribución normal

H1: Los residuos no siguen una distribución normal

α: 0,05

W = 0.91718, p-value = 0.03379

En este caso, el valor p es menor que 0.05 por lo tanto se rechaza H0, es decir los residuales no siguen una distribución normal.

Con respecto al test de Ljung-Box se obtuvo lo siguiente:

H0: Los residuales se distribuyen de forma independiente

H1: Los residuales no se distribuyen de forma independiente

α: 0.05

Q* = 8.0385, df = 6, p-value = 0.2353

Como se puede apreciar p-value mayor que 0.05, por lo tanto, se podría aceptar H0, es decir, los residuales se distribuyen independientemente. Sin embargo, los residuos no siguen

una distribución normal, es decir, este modelo quedaría descartado ya que sus residuos no serían un ruido blanco ni seguirían una distribución normal.

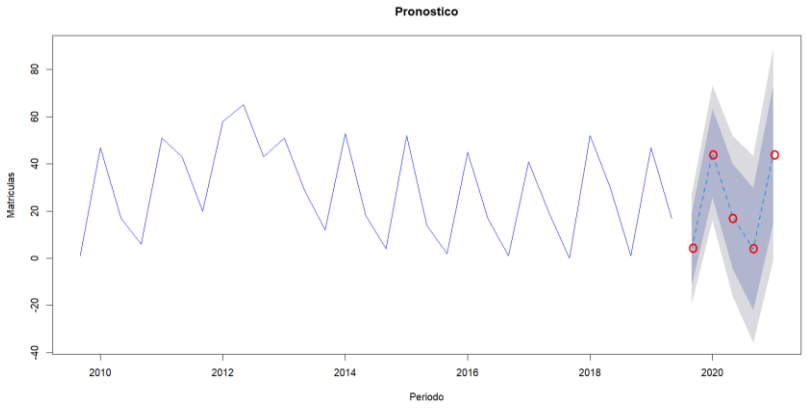
5.2.2.6. Pronósticos por el método HOLT-WINTERS

De acuerdo al análisis Se realizo la evaluación de pronóstico cinco periodos (2020-0 al 2021-1) como también para veintidós periodos (2021-2 al 2028-2) en la que se aplicara el método Holt-Winters.

A. Periodo 2020-0 al 2021-1

Teniendo en cuenta que el modelo se encuentra descartado por no superar los estadísticos de prueba, se procedió al cálculo del RMSE cuyo resultado fue 11.786661.

Figura 36 *Pronostico de cinco periodos (2020-0 al 2021-1) por el modelo Holt- winters para para el curso de Calculo III*



Fuente: elaboración propia

Cuadro 14 *Tabla de pronóstico para cinco periodos para del curso calculo III por el método Holt-Winters según ANEXO 5*

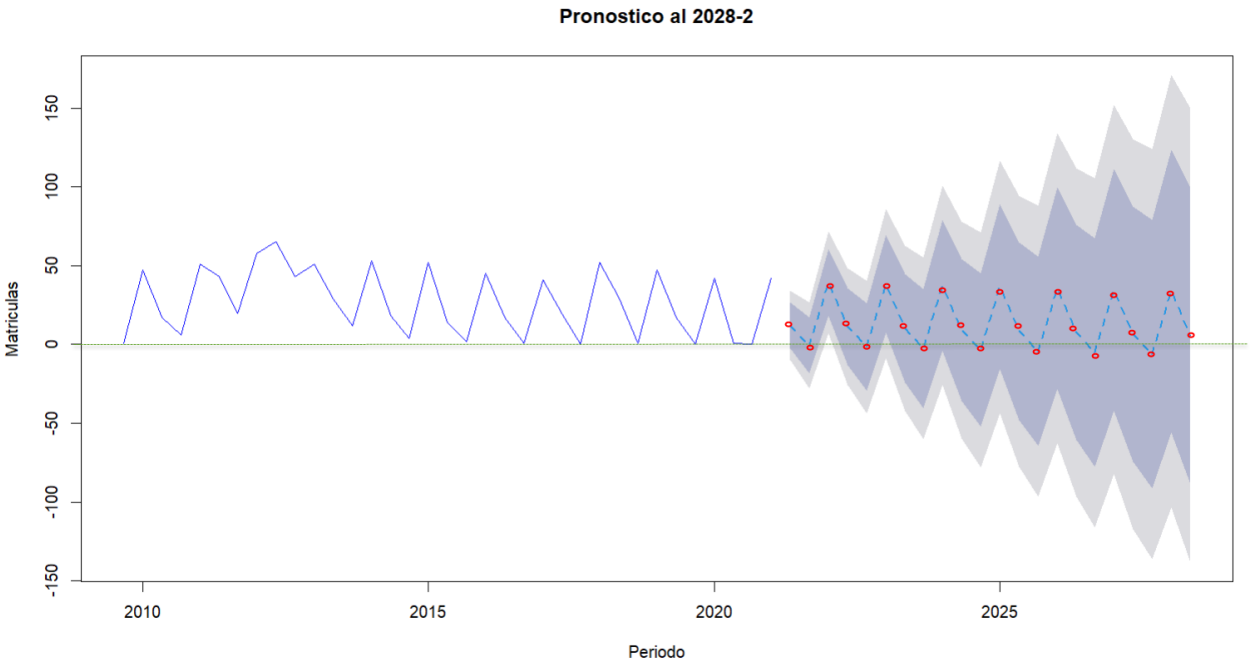
PERIODO	POINT_FORESCAST	MATRICULAS
2020-0	4.052566	4
2020-1	44.628311	44
2020-2	17.869963	17
2021-0	3.922784	3
2021-1	44.498528	44

El siguiente cuadro muestra los valores de matrículas que se daría con este método. Sin embargo, no se considera fiable el método.

B. Periodo 2021-2 al 2028-2

Aplicando las mismas pruebas de validación de los residuales que están dentro del intervalo de confianza ACF y los p-value para el test de Shapiro - Wilks se obtuvo 0.02709, no sigue una distribución normal. Para la prueba de Ljung-Box, se obtuvo 0.2546, se distribuyen de forma independiente. Es decir, este modelo también queda descartado para esta serie. Asimismo, se calculó el RMSE cuyo valor fue 11.260040.

Figura 37 *Pronostico por modelo Holt- winters para veintidós periodos (2021-2 al 2028-2) para el curso de Calculo III*



Fuente: elaboración propia

Cuadro 15 *Tabla de pronóstico por el método HOLT-WINTERS para veintidós periodos del curso calculo III según ANEXO 5*

PERIODO	POINT_FORESCAST	MATRICULAS
2021-2	12.0981151	12
2022-0	-0.6487846	0
2022-1	39.2538816	39
2022-2	11.1907371	11
2023-0	-1.5561627	0
2023-1	38.3465035	38
2023-2	10.283359	10
2024-0	-2.4635407	0

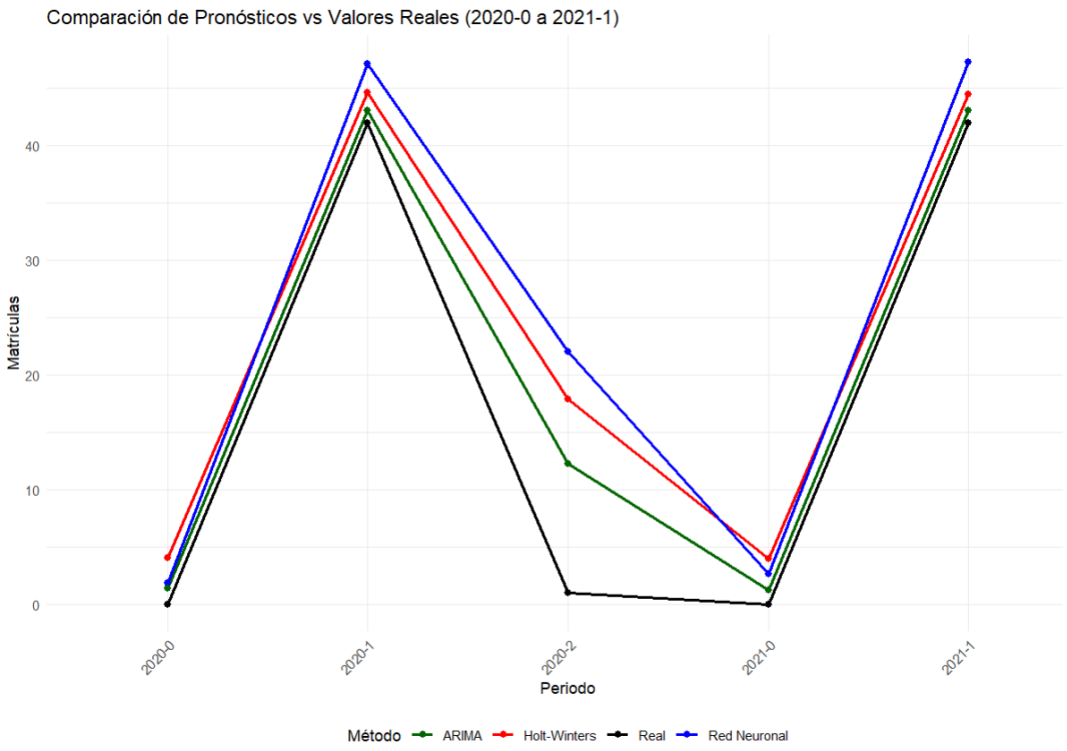
2024-1	37.4391255	37
2024-2	9.375981	9
2025-0	-3.3709187	0
2025-1	36.5317475	37
2025-2	8.468603	8
2026-0	-4.2782968	0
2026-1	35.6243694	36
2026-2	7.5612249	8
2027-0	-5.1856748	0
2027-1	34.7169914	35
2027-2	6.6538469	7
2028-0	-6.0930528	0
2028-1	33.8096134	34
2028-2	5.7464689	6

Tal como se advirtió la serie presenta valores negativos al igual que en el curso de Algoritmos y programación.

5.2.2.7. Selección del Modelo

Así pues, de la misma manera se presentarán los resultados para los pronósticos para el periodo 2020-0 al 2021-1, los cuales esta descritos y comparados por cada método en la figura 38. Posteriormente, se describen los resultados al pronosticar el periodo 2021-2 al 2028-2, tal como se muestra en la figura 39 para finalmente realizar la comparación de calidad de cada periodo por los tres métodos en el cuadro 16.

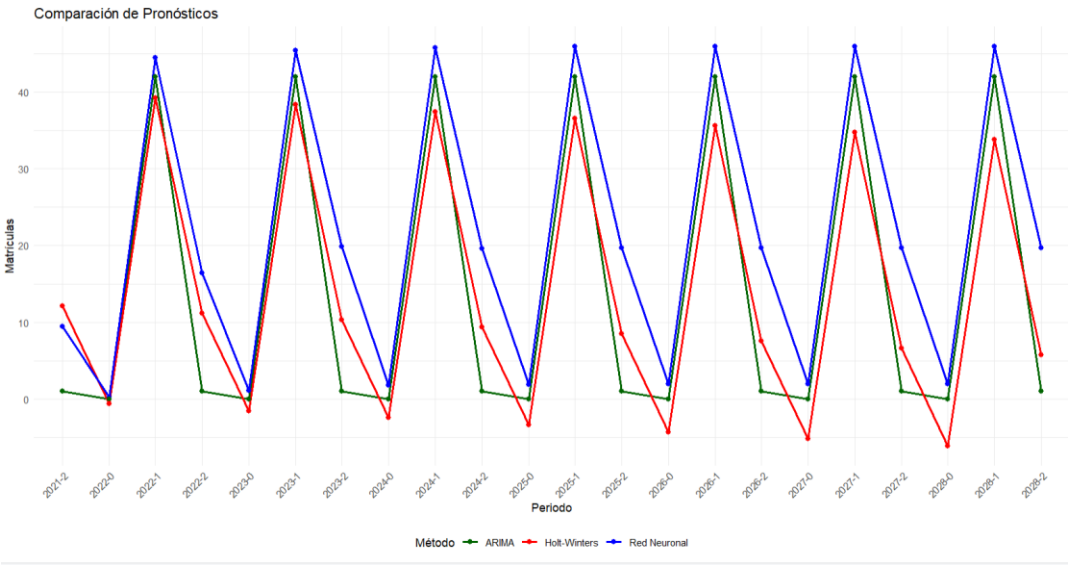
Figura 38 *Pronóstico 2020-0 al 2021-1 por los tres métodos del curso cálculo III*



Fuente: elaboración propia

Como se puede apreciar en la figura 38, presenta los datos reales y los pronósticos, notándose las diferencias y que tan alejadas se encuentran respecto a la serie de control. El pronóstico más cercano al real se visualiza es la de ARIMA en la que también se reflejará en la calidad de pronóstico indicado en el cuadro 16. Así pues, también se puede apreciar el alejamiento de los tres métodos en el rango de los periodos 20201- al 2021-0, esto debido al comportamiento en estos últimos periodos (trimestrales), siendo datos aleatorios para la serie que se veía sin tanta varianza antes de estos periodos.

Figura 39 *Pronóstico 2021-2 al 2028-2 por los tres métodos del curso de cálculo III*



Fuente: elaboración propia

Como se puede apreciar en la figura 39, los pronósticos son más estables en los periodos de verano es decir “ciclo-0” al siguiente “ciclo -1”, la varianza, alejamiento en valores se encuentran en los periodos “ciclo -1” al “ciclo -0” siguiente, esto también debido al comportamiento de la serie.

Cuadro 16 *Tabla de comparación de resultados RMSE de pronósticos del curso calculo III*

MODELO	RMSE (h=5)	RMSE (h=22)
Red Neuronal	8.090142	8.050362
Método ARIMA	5.138969	12.29406
Método HOLT-WINTERS	11.786661	11.260040

En conclusión, después de haber analizado cada uno de los pronósticos para los dos periodos se concluye que, para un pronóstico de cinco periodos el método ARIMA fue la de mejor

calidad. Sin embargo, para veintidós periodos el método de Redes Neuronales será la más adecuada para garantizar la calidad de pronóstico. Asimismo, el método Holt-Winters no pasó las pruebas en ninguno de los periodos por la cual se encuentra descartado.

5.3. Prueba de hipótesis

5.3.1. Hipótesis secundaria 1

A continuación, se enuncia la hipótesis a probar: el uso de redes neuronales aplicadas a series de tiempo incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y confecciones de la UNMSM. Como se puede apreciar en los numerales 5.2.1.1 y 5.2.2.1 de los cursos de Algoritmos y programación, así como Cálculo III, el modelo de redes neuronales, los residuales en ambos casos son cercanos a cero, además están dentro de los intervalos de confianza de ACF, asimismo siguen una distribución normal y están independientemente distribuidos, por lo tanto, se comprueba la hipótesis planteada. Además de tener el mejor RMSE como se puede apreciar en el numeral 5.2.1.7 en el curso de algoritmos y programación como también en el pronóstico para veintidós periodos del curso de cálculo III de acuerdo al numeral 5.2.2.7.

5.3.2. Hipótesis secundaria 2

Para la siguiente hipótesis enunciado: la aplicación de Holt-Winters incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y

Confecciones de la UNMSM. Según las observaciones del numeral 5.2.1.5 del curso de algoritmo y programación, donde se evidencia que los residuales son cercanos a cero es decir siguen una distribución normal, así mismo están dentro del intervalo de confianza ACF y están distribuidos independientemente.

De la misma manera siguiendo el análisis para el curso de cálculo III, según el numeral 5.2.2.5, donde se evidencia que los residuales no siguen una distribución normal, están dentro del intervalo de confianza ACF y están distribuidos independientemente. Por lo tanto, no cumple con todos los requisitos planteados, es decir, el modelo Holt- Winters no presenta evidencia a favor en los resultados de ambos cursos, solo de una y por lo tanto se estaría rechazando la hipótesis planteada.

5.3.3. Hipótesis secundaria 3

En la hipótesis a probar planteada: la aplicación de modelos ARIMA incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM. Como se puede apreciar en los numerales 5.2.1.3 y 5.2.2.3 de los cursos de Algoritmos y programación, así como Cálculo III, en el modelo ARIMA, los residuales en ambos casos son cercanos a cero es decir siguen una distribución normal, además están dentro de los intervalos de confianza de ACF y están distribuidos independientemente, por lo tanto, se comprueba la hipótesis planteada. Además, se indica que es el mejor modelo con mayor calidad para pronosticar para un periodo cinco del curso de cálculo III según los resultados indicados en el numeral 5.2.2.7, por tener el menor RMSE a comparación de los otros modelos.

5.3.4. Hipótesis general

En la hipótesis a probar planteada: el uso del análisis de series de tiempo incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM. Según la evaluación obtenida sobre las hipótesis secundarias que dos de los tres modelos se serie de tiempo analizado (ARIMA y Redes Neuronales) ambos determinaron resultados favorables de acuerdo a los criterios estadísticos que se propuso para evaluar la calidad del pronóstico. Para estos dos casos, los residuales son cercanos a cero, siguen una distribución normal, se encuentran dentro de los intervalos de confianza de la función de autocorrelación (ACF) y están distribuidos independientemente. Así también, los RMSE fueron los menores valores de error, la cual refuerzan la capacidad predictiva. Por otro lado, el modelo Holt-Winters no mostró resultados favorables para ambos cursos, la cual limita la su validez.

En conjunto, la evidencia estadística permite aceptar la hipótesis, debido a que se ha demostrado que el uso de modelos de series de tiempo, particularmente los de mayor robustes estadística como Redes Neuronales y ARIMA incide de manera positiva en la calidad del pronóstico de la demanda de cursos en la escuela profesional de ingeniería textil y confección de la UNMSM.

CAPÍTULO VI: DISCUSIÓN

Al realizar el tratamiento de la data de las matrículas de los cursos de Algoritmos y programación, Calculo III mediante los métodos Arima, Red Neuronal y Holt Winters se evaluaron si eran aptos para la investigación analizando tres estadísticos de prueba en sus residuales:

1) Intervalo de confianza ACF, los residuales son cercanos a cero y están dentro del intervalo de confianza ACF.
2) Prueba de Shapiro-Wilks, los residuales siguen una distribución normal, para un grado de significancia del 5%.
3) Prueba de ljung-box, los residuales se distribuyen independientemente, para un grado de significancia del 5%.

Así mismo, de cumplir los residuales con los tres estadísticos, se considera un método apto para su aplicación en la investigación teniendo en cuenta la calidad de pronóstico (el cual puede medirse mediante el indicador RMSE). En este sentido, se detalla los resultados al tratamiento de la data por los tres métodos.

La aplicación del método Red Neuronal se ajusta en ambos cursos, como también a los dos periodos pronosticados, debido a que cumplen con los tres estadísticos de prueba, concluyendo como un modelo fiable. Estos resultados concuerdan con lo establecido por Albán (2019), quien establece a las redes neuronales en base a los algoritmos perceptrón multicapa y función de base radial como algoritmos de alto rendimiento para predecir el abandono escolar

universitario tomando como datos utilizados de 2670 estudiantes matriculados en pregrado pertenecientes a las carreras de ciencia administrativas y humanas provenientes de universidades públicas del país de ecuador, correspondientes del primer año al cuarto en los periodos académicos del 2014 al 2017. La participación para el algoritmo perceptrón el 60% (1602 casos) de entrenamiento, el 30% (801 casos) de prueba y dejando un 10% (267 casos) para reserva perceptrón. Así mismo en caso del algoritmo multicapa el 70% (1869 casos) de entrenamiento y el 30% (801 casos) para la prueba. Así pues, para analizar la factibilidad de modelos trazo la curva ROC (características del funcionamiento del receptor) como también evaluó a los modelos y su capacidad para predecir. En conclusión, ambos algoritmos muestran rendimiento optimo en termino de precisión de predicción en caso del perceptrón multicapa resulta una tasa de predicción del 96.3%, mientras la tasa de predicción de la base de función radial corresponde al 96.8% lo que indica que los modelos de redes neuronales son fiables para la identificación de los estudiantes en riesgo de desertar de las universidades.

Así como también Ancco (2021), en su investigación donde compara dos métodos para mejorar el pronóstico de ventas de calzado mediante las redes neuronales y el modelo exponencial ETS; por la cual se tomaron como datos de ventas desde febrero del 2014 a marzo del 2021 como entrenamiento y de prueba entre abril del 2021 a junio del 2021. La ejecución del pronóstico se realizó en 8 pasos hacia adelante y se obtuvo que el modelo ETS presenta proyecciones de ventas sobre estimada sin embargo el modelo red neuronal presentan proyecciones que subestiman ligeramente las ventas de años anteriores y están más cercanos a las ventas reales. Por lo tanto, se concluye que el modelo de redes neuronales presenta mejores resultados en las evaluaciones de proyección siendo fiable para el pronóstico.

La aplicación del método ARIMA se ajusta en los cursos de algoritmos y programación como también en calculo III debido a que cumplen con los tres estadísticos de prueba en ambos periodos analizados, concluyendo como un modelo fiable. Estos resultados coinciden con la investigación de Botero y Cano (2008), en la cual se aplica el método ARIMA para el pronóstico del precio de la energía en la bolsa de Colombia por el cual se consideraron dos periodos de mayo del 1998 a julio del 2000 y junio del 2001 hasta junio del 2007. Este método se comparó con otro método referencia simple o ingenuo (SIMP) en la cual el estadístico MAPE = 1.92% resultó inferior a MSIMP=5.11%, con la cual se justifica implementar el modelo ARIMA el cual cumple con los estadísticos de prueba en la significancia de los parámetros, por lo tanto, es un modelo con buena capacidad de predicción.

La aplicación del método Holt-Winters no se ajusta en los cursos de algoritmos y programación y en el curso de cálculo III debido a que cumplen solo dos estadísticos de prueba (los residuales están dentro del Intervalo de confianza de ACF o siguen una distribución normal), sin embargo, los residuales no se distribuyen independientemente o no siguen una distribución normal. Si bien es cierto, la investigación de López (2021) también utiliza este método, en la cual busca el mejor pronostico entre el método Box-Jenkins y Holt-Winters utilizando datos del tráfico de pasajeros vía aérea en Chile en el periodo enero del 2015 a diciembre del 2019. En resumen, se obtuvo un RMSE para Box-Jenkins igual a 35.567 siendo un valor menor que el de Holt-Winters de 73.299 evidenciando que es menos fiable.

En la aplicación de los tres métodos podemos apreciar que, los residuales de los tres métodos son cercanos a cero además están dentro del intervalo de confianza AFC superando el primer estadístico de prueba. Para los estadísticos de distribución normal y distribución de forma independiente (prueba de Shapiro-Wilks y Ljung-Box), cumplen los métodos ARIMA y Redes Neuronales en ambos cursos; el método Holt-Winters solo superó la prueba de Shapiro-Wilks en el curso de Algoritmos y programación. Así mismo, al cumplir tres estadísticos, los métodos ARIMA y Redes Neuronales son aceptables para la aplicación a la presente investigación. Sin embargo, la evaluación será mediante la medición de la calidad RMSE.

Finalmente, se concluye mediante los estadísticos de prueba (los residuales están dentro del Intervalo de confianza de ACF, siguen una distribución normal y se distribuyen independientemente) los métodos a considerar son Redes Neuronales y ARIMA. Sin embargo, se tendrá que decidir entre ambos pronósticos el más cercano al real y dependerá de los periodos a los que se requiera evaluar pronosticar, considerando así el estadístico de error RMSE por cada curso y periodo.

Así pues, para el método Red Neuronal y ARIMA en el curso de Algoritmos y programación obtuvieron el RMSE más bajo para pronósticos para periodos de cinco y veintidós respectivamente, tomando el menor, el método seleccionado es redes Neuronales. Así mismo, para el curso de Calculo III el RMSE más bajo obtenido son para cinco periodos es el método ARIMA, sin embargo, para veintidós periodos es el método de Redes Neuronales, concluyendo que el método con mayor calidad de pronóstico serán cualquier de las dos, por lo que dependerá de los periodos a pronosticar, varianza, versatilidad de la serie, entre otros. Por lo contrario, el

método de Holt-Winters está más alejado del promedio al promedio, poco ajustable a valores positivos y tiene la más baja calidad de pronóstico, es decir el más alto RMSE cuando cumple los requisitos estadísticos o en otros casos no aplica.

Para concluir, al realizar los pronósticos de los cursos Calculo III, Algoritmos y Programación en los periodos 2020-0 al 2020-1 mediate los tres métodos se concluye que los métodos de Programación Neuronal y ARIMA son los más confiables debido a que pasaron los estadísticos de prueba, como también obtuvieron el menor coeficiente de RMSE. Así mismo, el modelo elegido varia en ambos cursos, ARIMA fue el de mejor calidad en el curso de Calculo III y Redes Neuronales en el curso de Algoritmos y Programación, esto debido a la notable estacionalidad que se presenta en los últimos periodos del curso de Calculo III, atribuyéndose mayor peso a los valores finales de la serie. En este sentido, el método ARIMA pose mayor ventaja frente para pronósticos cortos con serie de poca aleatoriedad.

Por otro lado, al realizar los pronósticos para ambos cursos para veintidós periodos el método con mejor calidad es el de redes neuronales, esto debido a la ventaja de aprendizaje que se obtiene del histórico de la serie y la evaluación de los residuos de cada uno de ellas.

CAPÍTULO VII: CONCLUSIONES Y RECOMENDACIONES

7.1. Conclusiones

1. Según lo establecido en la prueba de hipótesis, el uso del análisis de series de tiempo incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM. En este sentido, se aplicó la prueba de tres estadísticos a los dos cursos (los residuales deben estar dentro del intervalo de confianza ACF, tener una distribución normal y tener una distribución de forma independiente), en la cual se cumple en las redes neuronales y el método ARIMA, sin embargo, no se cumple para el método Holt-Winters quedando descartado. Así mismo, ambos métodos presentan bajos errores de pronóstico (RMSE).

Por otro lado, el estudio de series de tiempo enfocado a la calidad de pronóstico propicia una mejora en la planificación de demanda de cursos, en la cual se busca generar un escenario donde se disminuyan los cuellos de botella para mejorar los tiempos de procesos, optimizar los recursos administrativos, así como también mejorar en la planificación de docentes, optimizar la programación de horarios de cursos en apertura, entre otros. Por lo tanto, la aplicación de series de tiempo incide positivamente en la calidad de pronóstico en la demanda de cursos en la escuela profesional de ingeniería textil y confecciones de la UNMSM.

2. Tal como se establece en la prueba de hipótesis, el uso de redes neuronales aplicadas a series de tiempo incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y confecciones de la UNMSM. En este sentido el método de redes neuronales cumple con la prueba de los tres estadísticos, demostrando ser eficiente en el pronóstico de series de tiempo para los cursos de algoritmos y programación como calculo III. Sin embargo, al realizar pronósticos para veintidós periodos en el curso de cálculo III obtuvo una mejor calidad que el de ARIMA. Por lo tanto, el método de Redes neuronales se considera fiable para la investigación puesto obtuvo la mejor calidad al pronosticar veintidós y cinco periodos en el curso de Algoritmos y programación y al pronosticar veintidós periodos en el curso de Calculo III.

3. Según se establece en la prueba de hipótesis La aplicación de Holt-Winters no incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM. Debido a que el método supera los estadísticos de prueba para el pronóstico de cinco periodos, pero no para el pronóstico de veintidós periodos en el curso de algoritmos y programación, así mismo en el curso de cálculo III no se superan dos pruebas estadísticas (los residuales no siguen una distribución normal o no se distribuyen en forma independiente) para los pronósticos de ambos periodos. Por lo tanto, el método Holt-Winters no se considera fiable para la presente investigación debido a que no cumplen con todos los estadísticos de prueba descartándose para el presente estudio.

4. De acuerdo a lo establecido en la prueba de hipótesis, la aplicación del modelo ARIMA incide en la calidad del pronóstico de la demanda de cursos en la Escuela Profesional de Ingeniería Textil y Confecciones de la UNMSM. En este aspecto el método ARIMA superó los tres estadísticos de prueba por la que se considera un método confiable para el pronóstico, sin embargo, al realizar pronóstico de cinco periodos en el curso de Calculo III, este método presentó mejora calidad que el de redes neuronales puesto que la serie presentaba baja aleatoriedad y notable estacionalidad en los últimos valores. Por lo tanto, el método ARIMA es fiable para la investigación puesto que obtuvo la mejor calidad al realizar el pronóstico de cinco periodos en el curso de Calculo III.

7.2. Recomendaciones

1. En la actual coyuntura relacionada con el virus Covid-19 se propicia el estudio académico de forma virtual como también semi presencial generando nuevos escenarios en todos los sectores de una organización educativa el cual influye en los diversos pronósticos analizados. En este contexto se recomienda la aplicación de series de tiempo en los pronósticos de demanda de vacantes al momento de la apertura de un curso, puesto el problema se presenta en distintas otras escuelas profesionales. Así mismo, es necesario el cumplimiento de los tres estadísticos para

reconocer a un método confiable como también considerar el menor error de pronóstico MSE el cual permite hallar la mejor propuesta para

2. Al aplicar el método de las redes neuronales se recomienda tener muy en cuenta analizar también el método ARIMA puesto que este último podría tener un menor error de pronóstico al momento de proyectar las series de tiempo. Así mismo de tener una amplia data en la serie temporal, considerar un análisis multicapa puesto que el método neuronal nos ofrece una versatilidad al momento de aprender las series para proyectar con el menor error posible.

3. Se recomienda al usar el método Holt-Winters verificar los estadísticos de prueba a los residuales en las que tiene una distribución normal y siguen una distribución de forma independiente puesto que este método no superó las pruebas en la presente investigación y por antecedentes no sigue un patrón de confiabilidad. Así también, no recomendable para series con valores iguales o cercanos a cero, puesto que resultan valores negativos en su pronóstico y requiere de un ajuste externo para restringir valores positivos.

4. Dentro de las aplicaciones de series de tiempo se considera al método ARIMA como un método con mejor calidad y se adecuan para casi todas las series según los antecedentes. En este contexto se recomienda considerar este método para mejorar la calidad en los pronósticos, sin embargo, se debe tener en cuenta los datos muy atípicos, presentan diferentes patrones de inestabilidad (volatilidad), esto se debe a

que en la varianza de las observaciones cambia con el tiempo (heteroscedasticidad). Pues frente a estas altas varianzas densas se puede optar por otros métodos como las redes neuronales, por ejemplo.

5. Al momento de comparar los métodos ARIMA y Redes Neuronales, se tienen que considerar el comportamiento de la serie, en estacionalidad, varianza, ruido, como también la cantidad de periodos a pronosticar, puesto que este último puede definir la precisión y calidad del pronóstico.

6. Se recomienda continuar con la investigación aplicado a otras escuelas profesionales vinculados a ingeniería como también para otras escuelas profesionales segmentadas como son el caso de ciencias administrativas, contables, de la salud, entre otros. En este contexto se pretende hallar comportamientos cíclicos, predeterminados para la mejora en la calidad en pronósticos.

REFERENCIAS BIBLIOGRÁFICAS

Akiva, Y. (1955). orrelation theory of processes with random stationary n th increments. *Matematicheskii Sbornik, 79*, 141-196.

Albán, M. (2019). *Contribuciones a la predicción de la deserción universitaria a través de mineria de datos - Lima [Tesis de doctorado, UNMSM].* Repositorio universitario, Lima. Obtenido de https://hdl.handle.net/20.500.12672/10776

Ancco, T. (2021). *Análisis comparativo de series de tiempo para proyectar las ventas en las jerarquías de calzado en una empres del sector retail - Lima [tesis de licenciatura, UNMSM].* Repositorio Universitario, Lima. Obtenido de https://hdl.handle.net/20.500.12672/17282

Bishop, C. (2006). *Reconocimiento de patrones y aprendizaje automático.* New York: Springer-Verlag.

Botero, S., & Cano, J. (2008). Análisis de series de tiempo para la predicción de los precios de la energia en la bolsa de colombia. *Cuad. Econ. [online], 27*, 173-208.

Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2016). *Time Series Analysis.* Hoboken: John Wiley & Sons, Inc.

Callejas, I., Piñeros, J., Rocha, J., Hernández, F., & Delgado, F. (2013). Implementación de una red neuronal artificial tipo SOM en una FPGA para la resolución de trayectorias tipo laberinto. *II International Congress of Engineering Mechatronics and Automation (CIIMA).* Bogotá. doi:10.1109/ciima.2013.6682790

Castro, E., Castro, E., Rico, L., & Segovia, I. (1989). *Estimación en cálculo y medida.* Madrid: Síntesis, 1989.

Delgado, R. (2020). *Uso de los métodos multivariantes para el análisis del desempeño academico de los estudiantes de la educacion superior. (Caso: estudiantes integrantes en el primer curso de matematicas de la UNALM) - Lima [Tesis de Maestria, UNMSM].* repostitorio Universitario. Obtenido de https://hdl.handle.net/20.500.12672/14997

Fernandez, L. (2020). Ventajas de R como herramienta para el Análisis y Visualización de datos en Ciencias Sociales. *Revista Científica de la UCSA*, 97-111. doi:https://doi.org/10.18004/ucsa/2409-8752/2020.007.02.097

García, R. (2011). *Introducción a la probabilidad imposible: Estadística de la probabilidad o probabilidad estadística .* Madrid: 519 GARin.

Gestal, M. (2009). *Introducción a las redes de neuronas artificiales.* Departamento de Tecnologías de la Información y las Comunicaciones: Universidad da Coruña.

Gough, B. (2009). GNU scientific library reference manual. *Network Theory Ltd*.

Ihaka, R. (1998). R: Past and future history. *Computing Science and Statistics*.

Nava, M. (2017). Sinapsis. En *Actividades Integradoras del Aprendizaje por Sistemas, AIAS del sistema nervioso* (págs. 85-92). Colombia: Universidad del Rosario. doi:https://doi.org/10.2307/j.ctt1t6p6sm.14

Olabe, X. (1998). *Redes Neuronales Artificiales y sus Aplicaciones.* Publicaciones de la Escuela de Ingenieros.

Perez, L., Naranjo, M., Santos, O., Cabrera, J., & Nogueira, D. (2021). Algoritmos de minería de datos para la predicción del comportamiento de indicadores ambientales. *DYNA, 88(219), 12*, 228–236. doi:https://doi.org/10.15446/dyna.v88n219.95018

Reyes, A. (2020). *Conceptos Básicos de Análisis de Series de tiempo en el dominio del tiempo.* Madrid: Daniel José Reyes Valero.

Rivas, W., & Mazón, B. (2018). *Redes Neuronales Artificiales Aplicadas al Reconocimiento de Patrones.* Machala: UTMACH, 2018.

Rivera, E. (2015). Introduccion a las Redes Neuronales Artificiales. *Revista de investigaciones de la Universidad Don Bosco, 64*, 64. Obtenido de http://www.udb.edu.sv/editorial/en/publicaciones/detalle/cientifica_6

Rosenblatt, F. (1957). *Una automatización de percepción y reconocimiento, Laboratorio Aeronáutico de Cornell.* Laboratorio Aeronáutico de Cornell: Reporte 1957.

Rouhiainen, L. (2018). *Inteligencia Artificial.* Madrid: Alienta.

Rstudio. (09 de Mayo de 2023). *Rstudio*. Obtenido de https://posit.co/products/open-source/rstudio/

Sánchez, J., & Cuellar, M. (2018). *Análisis de series de tiempo con métodos econométricos para el control de gestion en redes de telecomunicaciones.* Bogotá: Fundación Universitaria San Mateo.

Santana, J., & Farfán, E. (2014). *El arte de programar en R: un lenguaje para la estadistica.* Jiutepec: Instituto Mexicano de Tecnología del Agua.

Terán, J., Ibarra, S., Laria, J., Castán, J., Treviño, M., Garcia, A., & Martínez, J. (2019). Estudio de redes neuronales para el pronóstico de la demanda. *Revista Facultad de Ingeniería*. doi:https://doi.org/10.19053/01211129.v28.n50.2019.8783

Uriel, E., & Muñiz, M. (1993). *Estadística económica y empresarial: teoría y ejercicios.* España: AC.

Vargas, L., & Fuquen, E. (2021). *Introducción al análisis de datos con RStudio.* Bogota: Estudio 45-8 S.A.S.

Velasquez, J., & Franco, C. (2012). Pronóstico de series de tiempo con tendencia y ciclo estacional usando el modelo airline y redes neuronales artificiales. *Ing. Cienc., 8*, 171–189.

Villareal, F. (2016). Introducción a los Modelos de Pronósticos., (págs. 1-121).

ANEXOS

ANEXO 1 – Matrículas del curso de Algoritmos y Programación en los periodos 2010-I al 2021-I.

Cuadro 17 *Tabla de matrículas en el curso de algoritmos y programación del periodo 2010-I al 2021-I*

Ciclo	Matriculas
2010-1	7
2010-2	46
2011-0	10
2011-1	55
2011-2	43
2012-0	21
2012-1	44
2012-2	45
2013-0	9
2013-1	56
2013-2	49
2014-0	37
2014-1	59
2014-2	38
2015-0	26
2015-1	80
2015-2	36
2016-0	10
2016-1	54
2016-2	46
2017-0	27
2017-1	69
2017-2	47

2018-0	4
2018-1	63
2018-2	48
2019-0	1
2019-1	64
2019-2	47
2020-0	15
2020-1	63
2020-2	40
2021-0	0
2021-1	49

Cuadro 18 *Tabla de matrículas en el curso de algoritmos y programación del periodo 2010-I al 2019-II utilizados para entrenamiento*

Ciclo	Matriculas
2010-1	7
2010-2	46
2011-0	10
2011-1	55
2011-2	43
2012-0	21
2012-1	44
2012-2	45
2013-0	9
2013-1	56
2013-2	49
2014-0	37
2014-1	59
2014-2	38
2015-0	26
2015-1	80

2015-2	36
2016-0	10
2016-1	54
2016-2	46
2017-0	27
2017-1	69
2017-2	47
2018-0	4
2018-1	63
2018-2	48
2019-0	1
2019-1	64
2019-2	47

Cuadro 19 *Tabla de matrículas en el curso de algoritmos y programación del periodo 2020-0 al 2021-I como data de prueba y control.*

Ciclo	Matriculas
2020-0	15
2020-1	63
2020-2	40
2021-0	0
2021-1	49

ANEXO 2 – Matrículas del curso de Calculo III en los periodos 2010-I al 2021-I.

Cuadro 20 *Tabla de matrículas en el curso de cálculo III del periodo 2010-I al 2021-I*

Ciclo	Matriculas
2010-0	1
2010-1	47

2010-2	17
2011-0	6
2011-1	51
2011-2	43
2012-0	20
2012-1	58
2012-2	65
2013-0	43
2013-1	51
2013-2	29
2014-0	12
2014-1	53
2014-2	18
2015-0	4
2015-1	52
2015-2	14
2016-0	2
2016-1	45
2016-2	17
2017-0	1
2017-1	41
2017-2	19
2018-0	0
2018-1	52
2018-2	30
2019-0	1
2019-1	47
2019-2	17
2020-0	0
2020-1	42
2020-2	1
2021-0	0
2021-1	42

Cuadro 21 *Tabla de matrículas en el curso de cálculo III del periodo 2010-0 al 2019-II utilizados como entrenamiento.*

Ciclo	Matriculas
2010-0	1
2010-1	47
2010-2	17
2011-0	6
2011-1	51
2011-2	43
2012-0	20
2012-1	58
2012-2	65
2013-0	43
2013-1	51
2013-2	29
2014-0	12
2014-1	53
2014-2	18
2015-0	4
2015-1	52
2015-2	14
2016-0	2
2016-1	45
2016-2	17
2017-0	1
2017-1	41
2017-2	19
2018-0	0
2018-1	52
2018-2	30
2019-0	1
2019-1	47
2019-2	17

Cuadro 22 *Tabla de matrículas en el curso de cálculo III del periodo 2020-0 al 2021-I como data de prueba y control.*

Ciclo	Matriculas
2020-0	0
2020-1	42
2020-2	1
2021-0	0
2021-1	42

ANEXO 3 – SCRIPT aplicado al curso de algoritmo y programación

Instalación de paquetes y librerías

```
install.packages("forecast")
install.packages("fpp2")
install.packages("ggplot2")
install.packages("tibble")
install.packages("dplyr")
install.packages("tidyr")
library(fpp2)
library(lavaan)
library(forecast)
library(ggplot2)
library(tibble)
library(dplyr)
library(tidyr)
```

Ingreso de datos para el análisis de la serie

Se ingresó los datos de los periodos 2010-I al 2021-I, (según la tabla 1 del anexo 1) posteriormente se plotea (grafica) la serie.

```
data= c(7,46,10,55,43,21,44,45,9,56,49,37,59,38,26,
      80,36,10,54,46,27,69,47,4,63,48,1,64,47,15,
      63,40,0,49)
#Conversion a serie
serie = ts(data, start = c(2010, 1), frequency = 3)
plot(serie, main="Serie Temporal", xlab="Tiempo", ylab="Serie")
```

```
fit=decompose(serie,type='additive')
plot(fit,xlab="Tiempo")
train = window(serie, end = c(2019, 2))
test = window(serie, start = c(2020, 0))
```

Aplicación de la Red Neuronal

Se ingresó los datos de los periodos de prueba (según la tabla 2 del anexo1)

```
#h=5 (train)
set.seed(123)
modelo_RN=nnetar(train)
pronostico_RN=forecast(modelo_RN,5)
pronostico_RN
plot(pronostico_RN, xlab="periodo",ylab="Matriculas")
checkresiduals(pronostico_RN)
shapiro.test(pronostico_RN$residuals)
accuracy(pronostico_RN,test)
#h=22 (serie)
modelo2_RN=nnetar(serie)
pronostico2_RN=forecast(modelo2_RN,22)
pronostico2_RN
plot(pronostico2_RN, xlab="periodo",ylab="Matriculas")
checkresiduals(pronostico2_RN)
shapiro.test(pronostico2_RN$residuals)
residuos_RN2 = residuals(pronostico2_RN)
rmse_RN = sqrt(mean(residuos_RN2^2, na.rm = TRUE))
rmse_RN
```

Aplicación del modelo ARIMA

```
#h=5 (train)
modelo_arima=auto.arima(train, seasonal=T, approximation=T)
pronostico_arima=forecast(modelo_arima,h=5)
pronostico_arima
plot(pronostico_arima,xlab="Tiempo",ylab="Serie")
checkresiduals(modelo_arima)
shapiro.test(pronostico_arima$residuals)
accuracy(pronostico_arima,test)
#h=22 (serie)
modelo2_arima=auto.arima(serie, seasonal=T, approximation=T)
pronostico2_arima=forecast(modelo2_arima,h=22)
pronostico2_arima
plot(pronostico2_arima,xlab="Tiempo",ylab="Serie")
```

```
checkresiduals(modelo2_arima)
shapiro.test(pronostico2_arima$residuals)
residuos_ARIMA2=residuals(pronostico2_arima)
rmse_ARIMA2=sqrt(mean(residuos_ARIMA2^2, na.rm = TRUE))
rmse_ARIMA2
```

Aplicación del modelo HOLT-WINTERS

```
#h=5 (train)
modelo_HW=HoltWinters(train,seasonal='additive')
pronostico_HW =forecast(modelo_HW, h = 5)
pronostico_HW
plot(pronostico_HW,xlab="Periodo",ylab="Matriculas",main="Pronostico",col="blue",flty=2)
checkresiduals(pronostico_HW)
shapiro.test(pronostico_HW$residuals)
accuracy(pronostico_HW,test)

#h=22 (serie)
modelo_HW2=HoltWinters(serie,seasonal='additive')
pronostico_HW2=forecast(modelo_HW2, h = 22)
pronostico_HW2
checkresiduals(pronostico_HW2)
shapiro.test(pronostico_HW2$residuals)
residuos_HW=residuals(pronostico_HW2)
rmse_HW2=sqrt(mean(residuos_HW^2, na.rm = TRUE))
rmse_HW2
```

Gráficas

```
# GRAFICA HOLT-WINTERS H=22
serie_c = ts(c(serie, rep(NA, length(pronostico_HW2$mean))),
                  start = start(serie), frequency = frequency(serie))
pronostico_f= ts(c(rep(NA, length(serie)), pronostico_HW2$mean),
                    start = start(serie), frequency = frequency(serie))
plot(serie_c, type = "l", lwd = 2, col = "blue",
    ylim = c(0, max(serie, pronostico_HW2$mean)),
    xlab = "Tiempo", ylab = "Valores",
    main = "Serie original y pronóstico Holt-Winters")
lines(pronostico_f, col = "darkgreen", lwd = 2, lty = 2)

#GRAFICA COMPARATIVA METODOS h=5#
serie_test = as.numeric(test)
periodos_grafica = c("2020-0", "2020-1", "2020-2", "2021-0", "2021-1")
```

```
#DATOS TEST
data_test =data.frame(
  Periodo = periodos_grafica,
  Valor = serie_test,
  Metodo = "Real")
#DATOS PRONOSTICOS
data_rn =data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_RN$mean),
  Metodo = "Red Neuronal")
data_arima = data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_arima$mean),
  Metodo = "ARIMA")
data_hw = data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_HW$mean),
  Metodo = "Holt-Winters")
#TODOS LOS DATOS
data_todos = rbind(data_test, data_rn, data_arima, data_hw)
# ESTRUCTURAR GRAFICO
ggplot(data_todos, aes(x = Periodo, y = Valor, color = Metodo, group = Metodo)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_color_manual(values = c("Real" = "black", "Red Neuronal" = "blue",
                                "ARIMA" = "darkgreen", "Holt-Winters" = "red")) +
  labs(
    title = "COMPARACION DE PRONOSTICOS Y DATOS DE CONTROL",
    x = "Periodo",
    y = "Matrículas",
    color = "Método" ) +
  theme_minimal() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 45, hjust = 1) )
```

```
#GRAFICA COMPARATIVA PARA h=22
periodos_grafica = c("2021-2", "2022-0", "2022-1", "2022-2", "2023-0", "2023-1", "2023-2",
"2024-0", "2024-1", "2024-2", "2025-0", "2025-1", "2025-2", "2026-0", "2026-1", "2026-2",
"2027-0","2027-1", "2027-2","2028-0","2028-1", "2028-2")
#DATOS PRONOSTICOS
data_rn =data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico2_RN$mean),
  Metodo = "Red Neuronal")
data_arima = data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico2_arima$mean),
  Metodo = "ARIMA")
data_hw = data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_HW2$mean),
  Metodo = "Holt-Winters")
#TODOS LOS DATOS
data_todos = rbind( data_rn, data_arima, data_hw)
# ESTRUCTURAR GRAFICO
ggplot(data_todos, aes(x = Periodo, y = Valor, color = Metodo, group = Metodo)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_color_manual(values = c( "Red Neuronal" = "blue",
                                 "ARIMA" = "darkgreen", "Holt-Winters" = "red")) +
  labs(
    title = "COMPARACION DE PRONOSTICOS POR METODO",
    x = "Periodo",
    y = "Matrículas",
    color = "Método" ) +
  theme_minimal() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 45, hjust = 1) )
```

ANEXO 4 – SCRIPT aplicados al curso de cálculo lll

Ingreso de datos para el análisis de la serie

Se ingresó los datos de los periodos 2010-0 al 2021-I, (según la tabla 4 del anexo 1) posteriormente se plotea (grafica) la serie.

```
data= c(1,47,17,6,51,43,20,58,65,43,51,29,12,53,18,
        4,52,14,2,45,17,1,41,19,0,52,30,1,47,17,
        0,42,1,0,42)
#CONVERSION A SERIE
serie = ts(data, start = c(2010, 0), frequency = 3)
plot(serie, main="Serie Temporal", xlab="Tiempo", ylab="Serie")
fit=decompose(serie,type='additive')
plot(fit,xlab="Tiempo")
train = window(serie, end = c(2019, 2))
test = window(serie, start = c(2020, 0))
```

Aplicación de la Red Neuronal

```
#h=5 (train)
set.seed(123)
modelo_RN=nnetar(train)
pronostico_RN=forecast(modelo_RN,5)
pronostico_RN
plot(pronostico_RN, xlab="periodo",ylab="Matriculas")
checkresiduals(pronostico_RN)
shapiro.test(pronostico_RN$residuals)
accuracy(pronostico_RN,test)

  #h=22 (serie)
modelo2_RN=nnetar(serie)
pronostico2_RN=forecast(modelo2_RN,22)
pronostico2_RN
plot(pronostico2_RN, xlab="periodo",ylab="Matriculas")
checkresiduals(pronostico_RN)
shapiro.test(pronostico_RN$residuals)
residuos_RN2 = residuals(pronostico2_RN)
rmse_RN = sqrt(mean(residuos_RN2^2, na.rm = TRUE))
rmse_RN
```

Aplicación del modelo ARIMA

```
#  ARIMA - h=5
log_train = log1p(train)
# Ajustar modelo ARIMA con estacionalidad automáticamente
modelo_arima = auto.arima(log_train, seasonal = TRUE)
summary(modelo_arima) #muestras el modelo
pronostico_log = forecast(modelo_arima, h = 5)
checkresiduals(modelo_arima)
shapiro.test(residuals(modelo_arima))
# Revertir transformación logarítmica solo en partes numéricas
pronostico_arima =pronostico_log
pronostico_arima$mean  =expm1(pronostico_log$mean)
pronostico_arima$lower = expm1(pronostico_log$lower)
pronostico_arima$upper = expm1(pronostico_log$upper)
print(pronostico_arima)
  pred_arima =pronostico_arima$mean
accuracy(pred_arima,test)

#  ARIMA - h=22
log_serie = log1p(serie)
modelo_arima2 = auto.arima(log_serie, seasonal = TRUE)
summary(modelo_arima2)
pronostico_log2 <- forecast(modelo_arima2, h = 22)
checkresiduals(modelo_arima2)
shapiro.test(residuals(modelo_arima2))
# Revertir transformación logarítmica solo en partes numéricas
pronostico_arima2 <- pronostico_log2
pronostico_arima2$mean  <- expm1(pronostico_log2$mean)
pronostico_arima2$lower <- expm1(pronostico_log2$lower)
pronostico_arima2$upper <- expm1(pronostico_log2$upper)
print(pronostico_arima2)
```

Graficas para el método ARIMA

```
  # Gráfico con datos Serie y pronóstico h=5
  autoplot(train, series = "Serie",color ="black") +
    autolayer(pronostico_arima$mean, series = "Pronóstico", color = "blue") +
    autolayer(pronostico_arima$lower[, 2], series = "Límite inferior 95%", linetype = "dashed",
color = "gray50") +
    autolayer(pronostico_arima$upper[, 2], series = "Límite superior 95%", linetype = "dashed",
color = "gray50") +
    ggtitle("Matrículas: Datos Históricos y Pronóstico ARIMA (5 periodos)") +
    xlab("Periodo") + ylab("Matrículas") +
    guides(color = guide_legend(title = "Series")) +
```

```
    theme_minimal()


# Gráfico pronóstico h=22
autoplot(serie, series = "Serie",color ="black") +
  autolayer(pronostico_arima2$mean, series = "Pronóstico", color = "blue") +
  autolayer(pronostico_arima2$lower[, 2], series = "Límite inferior 95%", linetype = "dashed",
color = "gray50") +
  autolayer(pronostico_arima2$upper[, 2], series = "Límite superior 95%", linetype = "dashed",
color = "gray50") +
  ggtitle("Matrículas: Datos Históricos y Pronóstico ARIMA (22 periodos)") +
  xlab("Periodo") + ylab("Matrículas") +
  guides(color = guide_legend(title = "Series")) +
  theme_minimal()
```

Aplicación del modelo HOLT-WINTERS

```
# Pronostico h=5
modelo_HW=HoltWinters(train,seasonal='additive')
pronostico_HW <- forecast(modelo_HW, h = 5)
pronostico_HW
plot(pronostico_HW,xlab="Periodo",ylab="Matriculas",main="Pronostico",col="blue",flty=2)
checkresiduals(pronostico_HW)
shapiro.test(pronostico_HW$residuals)
accuracy(pronostico_HW,test)

# Pronostico h=22
modelo_HW2=HoltWinters(serie,seasonal='additive')
pronostico_HW2 <- forecast(modelo_HW2, h = 22)
pronostico_HW2
checkresiduals(pronostico_HW2)
shapiro.test(pronostico_HW2$residuals)
plot(pronostico_HW2,xlab="Periodo",ylab="Matriculas",main="Pronostico al 2028-
2",col="blue",flty=2)
residuos_HW2=residuals(pronostico_HW2)
rmse_HW2=sqrt(mean(residuos_HW2^2, na.rm = TRUE))
rmse_HW2
```

Graficas comparativas

```
#ESTRUCTURAR DATOS PARA LA GRAFICA h=5#
serie_test = as.numeric(test)
periodos_grafica = c("2020-0", "2020-1", "2020-2", "2021-0", "2021-1")
#DATOS TEST
data_test =data.frame(
  Periodo = periodos_grafica,
  Valor = serie_test,
  Metodo = "Real")
#DATOS PRONOSTICOS
data_rn =data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_RN$mean),
  Metodo = "Red Neuronal")
data_arima = data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_arima$mean),
  Metodo = "ARIMA")
data_hw = data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_HW$mean),
  Metodo = "Holt-Winters")
#TODOS LOS DATOS
data_todos = rbind(data_test, data_rn, data_arima, data_hw)
# ESTRUCTURAR GRAFICO
ggplot(data_todos, aes(x = Periodo, y = Valor, color = Metodo, group = Metodo)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_color_manual(values = c("Real" = "black", "Red Neuronal" = "blue",
                                "ARIMA" = "darkgreen", "Holt-Winters" = "red")) +
  labs(
    title = "Comparación de Pronósticos vs Valores Reales (2020-0 a 2021-1)",
    x = "Periodo",
    y = "Matrículas",
    color = "Método" ) +
  theme_minimal() +
  theme(legend.position = "bottom",
    axis.text.x = element_text(angle = 45, hjust = 1) )


#ESTRUCTURAR DATOS PARA LA GRAFICA h =22#
periodos_grafica = c("2021-2", "2022-0", "2022-1", "2022-2", "2023-0", "2023-1", "2023-2", "2024-
0",  "2024-1",  "2024-2",  "2025-0",  "2025-1",  "2025-2",  "2026-0",  "2026-1",  "2026-2",    "2027-
0","2027-1", "2027-2","2028-0","2028-1", "2028-2")
#DATOS PRONOSTICOS
data_rn2 =data.frame(
```

```
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico2_RN$mean),
  Metodo = "Red Neuronal")
data_arima2 = data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_arima2$mean),
  Metodo = "ARIMA")
data_hw2 = data.frame(
  Periodo = periodos_grafica,
  Valor = as.numeric(pronostico_HW2$mean),
  Metodo = "Holt-Winters")
#TODOS LOS DATOS
data_todos2 = rbind(data_rn2, data_arima2, data_hw2)
# ESTRUCTURAR GRAFICO
ggplot(data_todos2, aes(x = Periodo, y = Valor, color = Metodo, group = Metodo)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_color_manual(values = c("Real" = "black", "Red Neuronal" = "blue",
                                "ARIMA" = "darkgreen", "Holt-Winters" = "red")) +
  labs(
    title = "Comparación de Pronósticos",
    x = "Periodo",
    y = "Matrículas",
    color = "Método" ) +
  theme_minimal() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 45, hjust = 1) )
```

ANEXO 5 – PRONOSTICOS del curso de cálculo III

A continuación, se presenta los resultados al pronosticar por los tres métodos en Rstudio, de acuerdo al SCRIPT del ANEXO4.

Figura 40 *Pronostico para cinco periodos por Redes Neuronales calculo III tomados de Rstudio*

```
> pronostico_RN
        Point Forecast
2019.67       1.838517
2020.00      47.184502
2020.33      22.014653
2020.67       2.659155
2021.00      47.303200
```

Fuente: elaboración propia

Figura 41 Pronostico para veintidós periodos por Redes Neuronales para el curso calculo III tomado de Rstudio

```
> pronostico2_RN
        Point Forecast
2021.33      12.524251
2021.67       0.798022
2022.00      43.913345
2022.33      16.956572
2022.67       1.329073
2023.00      45.091502
2023.33      17.789332
2023.67       1.498055
2024.00      45.617264
2024.33      18.124345
2024.67       1.563592
2025.00      45.813646
2025.33      18.229422
2025.67       1.586311
2026.00      45.881626
2026.33      18.270051
2026.67       1.594783
2027.00      45.904569
2027.33      18.281229
2027.67       1.597435
2028.00      45.912227
2028.33      18.286090
```

Fuente: elaboración propia

Figura 42 *Pronostico para cinco periodos por el método ARIMA calculo III tomados de Rstudio*

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2019.667	1.407766	0.238964	3.679183	-0.1284214	5.651539
2020.000	43.041724	20.226564	90.379529	13.4237500	133.477749
2020.333	12.244673	5.383467	26.480577	3.3376561	39.441510
2020.667	1.204221	-0.331613	6.269128	-0.6446183	12.671470
2021.000	43.041724	11.391633	155.530898	5.3326668	305.296463

Fuente: elaboración propia

Figura 43 *Pronostico para veintidós periodos por el método ARIMA para calculo III tomados de Rstudio*

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021.333	1	-0.1669108	3.801407	-0.4759744	6.633215
2021.667	0	-0.5834554	1.400703	-0.7379872	2.816608
2022.000	42	16.9114182	102.230240	10.2665499	163.114127
2022.333	1	-0.4203709	5.900965	-0.6991058	12.293711
2022.667	0	-0.7101855	2.450483	-0.8495529	5.646855
2023.000	42	11.4620250	147.370750	5.4692246	284.814778
2023.333	1	-0.5612022	8.115816	-0.8034220	19.348153
2023.667	0	-0.7806011	3.557908	-0.9017110	9.174077
2024.000	42	8.4341523	194.990052	3.2264278	436.485293
2024.333	1	-0.6529812	10.526752	-0.8626986	28.132987
2024.667	0	-0.8264906	4.763376	-0.9313493	13.566494
2025.000	42	6.4609047	246.825172	1.9519802	625.359223
2025.333	1	-0.7177933	13.174009	-0.8999172	38.966899
2025.667	0	-0.8588967	6.087004	-0.9499586	18.983449
2026.000	42	5.0674437	303.741186	1.1517807	858.288323
2026.333	1	-0.7659036	16.086976	-0.9248002	52.191605
2026.667	0	-0.8829518	7.543488	-0.9624001	25.595802
2027.000	42	4.0330732	366.369979	0.6167965	1142.619498
2027.333	1	-0.8028717	19.291354	-0.9421829	68.183739
2027.667	0	-0.9014359	9.145677	-0.9710915	33.591870
2028.000	42	3.2382583	435.264113	0.2430667	1486.450396
2028.333	1	-0.8320151	22.811660	-0.9547313	87.361371

Fuente: elaboración propia

Figura 44 *Pronostico para cinco periodos por método Holt-Winters calculo III tomados de Rstudio*

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2019.667	4.052566	-11.078762	19.18390	-19.0888038	27.19394
2020.000	44.628311	25.900954	63.35567	15.9872897	73.26933
2020.333	17.869963	-4.364942	40.10487	-16.1353892	51.87532
2020.667	3.922784	-21.976436	29.82200	-35.6866545	43.53222
2021.000	44.498528	15.109290	73.88777	-0.4484331	89.44549

Fuente: elaboración propia

Figura 45 Pronostico *para veintidós periodos por método Holt-Winters calculo III tomados de Rstudio*

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021.667	12.0981151	-2.345120	26.54135	-9.990907	34.18714
2022.000	-0.6487846	-18.393164	17.09559	-27.786471	26.48890
2022.333	39.2538816	18.256415	60.25135	7.141027	71.36674
2022.667	11.1907371	-13.078245	35.45972	-25.925468	48.30694
2023.000	-1.5561627	-29.117420	26.00509	-43.707468	40.59514
2023.333	38.3465035	7.441869	69.25114	-8.918056	85.61106
2023.667	10.2833590	-24.036084	44.60280	-42.203700	62.77042
2024.000	-2.4635407	-40.252764	35.32568	-60.257171	55.33009
2024.333	37.4391255	-3.890285	78.76854	-25.768754	100.64701
2024.667	9.3759810	-35.575819	54.32778	-59.371864	78.12383
2025.000	-3.3709187	-52.008412	45.26657	-77.755545	71.01371
2025.333	36.5317475	-15.865773	88.92927	-43.603345	116.66684
2025.667	8.4686030	-47.772298	64.70950	-77.544432	94.48164
2026.000	-4.2782968	-64.427270	55.87068	-96.268212	87.71162
2026.333	35.6243694	-28.506524	99.75526	-62.455368	133.70411
2026.667	7.5612249	-60.633227	75.75568	-96.733189	111.85564
2027.000	-5.1856748	-77.507457	67.13611	-115.792296	105.42095
2027.333	34.7169914	-41.804046	111.23803	-82.311836	151.74582
2027.667	6.6538469	-74.145487	87.45318	-116.918071	130.22576
2028.000	-6.0930528	-91.232731	79.04662	-136.302953	124.11685
2028.333	33.8096134	-55.739973	123.35920	-103.144661	170.76389
2028.667	5.7464689	-88.289277	99.78221	-138.068794	149.56173

Fuente: elaboración propia

ANEXO 6 – PRONOSTICOS del curso Algoritmos y programación

A continuación, se presenta los resultados al pronosticar por los tres métodos en Rstudio, de acuerdo al SCRIPT del ANEXO3.

Figura 46 *Pronostico para cinco periodos para el curso de Algoritmos y programación por el método Redes Neuronales tomados de Rstudio*.

	Point Forecast
2019.67	20.55220
2020.00	60.90157
2020.33	44.63939
2020.67	15.94115
2021.00	62.19591

Fuente: elaboración propia

Figura 47 *Pronostico para veintidós periodos para el curso de Algoritmos y programación por el método Redes Neuronales tomados de Rstudio,*

	Point Forecast
2021.33	43.05464
2021.67	16.80610
2022.00	57.66019
2022.33	43.39104
2022.67	15.08055
2023.00	57.65418
2023.33	43.41832
2023.67	15.19901
2024.00	57.64521
2024.33	43.42137
2024.67	15.19039
2025.00	57.64628
2025.33	43.42171
2025.67	15.19101
2026.00	57.64618
2026.33	43.42175
2026.67	15.19097
2027.00	57.64619
2027.33	43.42176
2027.67	15.19097
2028.00	57.64619
2028.33	43.42176

Fuente: elaboración propia

Figura 48 *Pronostico para cinco periodos para el curso de Algoritmos y programación por el método ARIMA tomados de Rstudio*

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2019.667	2.753632	-14.53158	20.03884	-23.68182	29.18908
2020.000	63.415456	46.13024	80.70067	36.98001	89.85091
2020.333	47.584544	30.29933	64.86976	21.14909	74.01999
2020.667	1.728557	-16.98905	20.44616	-26.89755	30.35467
2021.000	63.757148	45.03954	82.47475	35.13104	92.38326

Fuente: elaboración propia

Figura 49 *Pronostico para veintidós periodos para el curso de Algoritmos y programación por el método ARIMA tomados de Rstudio*

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021.333	44.947180	28.324232	61.57013	19.5245744	70.36979
2021.667	8.756254	-7.866694	25.37920	-16.6663515	34.17886
2022.000	58.831512	42.208564	75.45446	33.4089066	84.25412
2022.333	42.438694	25.052307	59.82508	15.8485091	69.02888
2022.667	4.316364	-13.070024	21.70275	-22.2738216	30.90655
2023.000	53.846408	36.460021	71.23280	27.2562230	80.43659
2023.333	43.710631	23.168438	64.25282	12.2940573	75.12720
2023.667	6.567626	-13.974567	27.10982	-24.8489477	37.98420
2024.000	56.374123	35.831930	76.91632	24.9575494	87.79070
2024.333	43.065691	21.029112	65.10227	9.3636520	76.76773
2024.667	5.426115	-16.610463	27.46269	-28.2759236	39.12815
2025.000	55.092436	33.055857	77.12901	21.3903972	88.79448
2025.333	43.392710	19.403693	67.38173	6.7046749	80.08075
2025.667	6.004922	-17.984094	29.99394	-30.6831127	42.69296
2026.000	55.742320	31.753303	79.73134	19.0542850	92.43036
2026.333	43.226894	17.703003	68.75079	4.1914708	82.26232
2026.667	5.711436	-19.812455	31.23533	-33.3239869	44.74686
2027.000	55.412794	29.888903	80.93669	16.3773711	94.44822
2027.333	43.310972	16.212064	70.40988	1.8667684	84.75517
2027.667	5.860249	-21.238658	32.95916	-35.5839538	47.30445
2028.000	55.579882	28.480974	82.67879	14.1356783	97.02408
2028.333	43.268340	14.742845	71.79383	-0.3576404	86.89432

Fuente: elaboración propia

Figura 50 *Pronostico para cinco periodos para el curso de Algoritmos y programación por el método HOLT-WINTERS tomados de Rstudio*

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2019.667	10.74290	-6.020459	27.50625	-14.89444	36.38024
2020.000	59.89706	42.775917	77.01821	33.71253	86.08160
2020.333	43.82668	25.926635	61.72672	16.45092	71.20243
2020.667	10.31806	-10.180782	30.81691	-21.03222	41.66834
2021.000	59.47223	37.201740	81.74272	25.41246	93.53200

Fuente: elaboración propia

Figura 51 *Pronostico para veintidós periodos para el curso de Algoritmos y programación por el método HOLT-WINTERS tomados de Rstudio*

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021.333	37.827639	22.100371	53.55491	13.7748586	61.88042
2021.667	1.704127	-14.399966	17.80822	-22.9249587	26.33321
2022.000	48.211559	31.290261	65.13286	22.3326664	74.09045
2022.333	30.492766	11.127700	49.85783	0.8764541	60.10908
2022.667	-5.630746	-26.843332	15.58184	-38.0725962	26.81110
2023.000	40.876685	17.256191	64.49718	4.7522562	77.00111
2023.333	23.157892	-4.616785	50.93257	-19.3198109	65.63560
2023.667	-12.965619	-44.003514	18.07227	-60.4339822	34.50274
2024.000	33.541812	-1.189259	68.27288	-19.5747780	86.65840
2024.333	15.823019	-24.158961	55.80500	-45.3241432	76.97018
2024.667	-20.300493	-64.588234	23.98725	-88.0327485	47.43176
2025.000	26.206939	-22.713609	75.12749	-48.6105822	101.02446
2025.333	8.488145	-46.451047	63.42734	-75.5340987	92.51039
2025.667	-27.635366	-87.685881	32.41515	-119.4747028	64.20397
2026.000	18.872065	-46.555579	84.29971	-81.1908818	118.93501
2026.333	1.153272	-70.906648	73.21319	-109.0528648	111.35941
2026.667	-34.970240	-112.811505	42.87103	-154.0181813	84.07770
2027.000	11.537192	-72.312962	95.38735	-116.7005526	139.77494
2027.333	-6.181601	-97.195013	84.83181	-145.3746035	133.01140
2027.667	-42.305113	-139.685449	55.07522	-191.2354846	106.62526
2028.000	4.202318	-99.746544	108.15118	-154.7737479	163.17838
2028.333	-13.516475	-125.108021	98.07507	-184.1810180	157.14807

>

Fuente: elaboración propia