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

Influencia de la regresión lineal, polinómica y redes neuronales en el modelamiento y pronóstico del Producto Bruto Interno (PBI) en función de la población

TESIS

Para optar el Título Profesional de Ingeniero Industrial

AUTOR

Miguel Israel CANCHUMANI MUÑOZ

ASESOR

Dr. 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

Canchumani, M. (2025). *Influencia de la regresión lineal, polinómica y redes neuronales en el modelamiento y pronóstico del Producto Bruto Interno (PBI) en función de la población*. [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	Miguel Israel Canchumani Muñoz
Tipo de documento de identidad	DNI
Número de documento de identidad	76027072
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	Francisco Javier Wong Cabanillas
Tipo de documento	DNI
Número de documento de identidad	07564428
Miembro del jurado 1	
Nombres y apellidos	Augusto Denis Madueño Macedo
Tipo de documento	DNI
Número de documento de identidad	07362033
Miembro del jurado 2	

Nombres y apellidos	Daniel Víctor Surco Salinas
Tipo de documento	DNI
Número de documento de identidad	09722150
Datos de investigación	
Línea de investigación	ODS 9: Industria, innovación e infraestructura 4. Gestión organizacional sostenible
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: Av. Carlos Germán Amezaga 375 Latitud: -12.05678° Longitud: -77.085105°
Año o rango de años en que se realizó la investigación	Julio 2022 - Abril 2024
URL de disciplinas OCDE	Estadística, Probabilidad https://purl.org/pe-repo/ocde/ford#1.01.03 Econometría https://purl.org/pe-repo/ocde/ford#5.02.02 Demografía https://purl.org/pe-repo/ocde/ford#5.04.02

VICEDECANATO ACADÉMICO

ACTA DE SUSTENTACIÓN N°003-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 **viernes 24 de enero de 2025**, a las 12:00 horas, dio inicio a la sustentación de la tesis:

"INFLUENCIA DE LA REGRESIÓN LINEAL, POLINÓMICA Y REDES NEURONALES EN EL MODELAMIENTO Y PRONÓSTICO DEL PRODUCTO BRUTO INTERNO (PBI) EN FUNCIÓN DE LA POBLACIÓN"

Que presenta el Bachiller:

MIGUEL ISRAEL CANCHUMANI MUÑOZ

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

Luego de la exposición, absueltas las preguntas del Jurado y siendo las 12:55 horas se procedió a la evaluación secreta, habiendo sido APROBADO por UNANIMIDAD con la calificación promedio de 16 (DIECISEIS), lo cual se comunicó públicamente.

Lima, 24 de enero del 2025

DR. FRANCISCO JAVIER WONG CABANILLAS
Presidente

DR. DANIEL VÍCTOR SURCO SALINAS
Miembro

MG. AUGUSTO DENIS MADUEÑO MACEDO
Miembro

DR. ALFONSO RAMÓN CHUNG PINZÁS
Asesor

MG. LUIS ROLANDO RAEZ GUEVARA
Vicedecano Académico – FII

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Facultad de Ingeniería Industrial
VICEDECANO ACADÉMICO

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 la Resolución Decanal **N°000093-2025-D-FII** de la tesis, cuyo título es **INFLUENCIA DE LA REGRESIÓN LINEAL, POLINÓMICA Y REDES NEURONALES EN EL MODELAMIENTO Y PRONÓSTICO DEL PRODUCTO BRUTO INTERNO (PBI) EN FUNCIÓN DE LA POBLACIÓN**, presentado por el bachiller **CANCHUMANI MUÑOZ MIGUEL ISRAEL** 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 **19%** 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, 03 de febrero del 2025

UNMSM
Firmado digitalmente por CHUNG PINZAS Alfonso Ramon FAU 20148092282 soft
Motivo: Soy el autor del documento
Fecha: 05.02.2025 09:36:39 -05:00

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

A María Genoveva Paima Córdova

ÍNDICE GENERAL

RESUMEN vii
ABSTRACT viii
CAPÍTULO I: INTRODUCCIÓN 1
1.1. Introducción 1
1.2. 5 2
1.2.1. Determinación del problema 2
1.2.2. Formulación del problema 3
1.2.2.1. Problema Principal 3
1.2.2.2. Problemas secundarios 3
1.3. Objetivos 4
1.3.1. Objetivo principal 4
1.3.2. Objetivos secundarios 4
1.4. Importancia y alcance de la investigación 4
1.5. Limitaciones de la investigación 5
CAPÍTULO II: REVISIÓN DE LITERATURA 6
2.1. Marco teórico 6
2.2. Antecedentes del estudio 6
2.3. Bases teóricas 10
2.3.1. Producto Bruto Interno (PBI) 10
2.3.2. Análisis de regresión 14
2.3.2.1. Regresión lineal simple: 14
2.3.2.2. Regresión lineal múltiple: 15
2.3.3. Regresión polinomial 19
2.3.4. Red neuronal artificial 20
2.4. Definición de términos 36

CAPÍTULO III: HIPÓTESIS Y VARIABLES 39
3.1. Hipótesis 39
3.1.1. Hipótesis principal 39
3.1.2. Hipótesis secundarias 39
3.2. Variables 39
3.3. Operacionalización de las variables 40
CAPÍTULO IV: MATERIALES Y MÉTODOS 46
4.1. Área de estudio 46
4.2. Tipo y diseño de investigación 46
4.3. Población y muestra 47
4.4. Procedimientos, técnicas e instrumentos de recolección de datos 47
4.5. Análisis estadístico 48
CAPÍTULO V: RESULTADOS 50
5.1. Preparación de datos 50
5.2. Modelo Lineal 54
5.3. Modelo Polinómico 58
Orden del modelo polinómico 59
5.4. Red Neuronal 61
5.5. Selección del mejor modelo 62
5.6. Pruebas e hipótesis 62
5.7. Script utilizado 63
CAPÍTULO VI: DISCUSIÓN 69
CAPÍTULO VII: CONCLUSIONES Y RECOMENDACIONES 72
7.1. Conclusiones 72
7.2. Recomendaciones 72
REFERENCIAS BIBLIOGRÁFICAS 74

INDICE DE TABLAS

Tabla 1 Métodos de cálculo del PBI 11
Tabla 2 Identificación de variables 39
Tabla 3 Definición operacional 41
Tabla 4 Análisis de los resultados 42
Tabla 5 Matriz de consistencia 43
Tabla 6 Áreas del conocimiento abordadas 46
Tabla 7 Base de datos 50
Tabla 8 Base de datos de entrenamiento 53
Tabla 9 Base de datos de prueba 54
Tabla 10 Resultados del modelo lineal 54
Tabla 11 Resultados de los modelos de orden 2 al 6 59
Tabla 12 Modelo polinómico de orden 7 60
Tabla 13 Resumen de verificación de supuestos 60
Tabla 14 Comparación de MAPE entre modelos 62

INDICE DE FIGURAS

Figura 1 Representación gráfica del error 16
Figura 2 Red neuronal artificial simple 20
Figura 3 Red neuronal compleja 21
Figura 4 Partes de la neurona 22
Figura 5 Principales funciones de activación 29
Figura 6 Gráfico PBI vs Población 52
Figura 7 Red neuronal propuesta como modelo rival 61

RESUMEN

La presente tesis busca, mediante la aplicación de tres métodos estadísticos: regresión lineal, regresión polinómica y las redes neuronales, evaluar la influencia que estos tienen al desarrollar un modelo que permita el pronóstico del Producto Bruto Interno (PBI) del Perú en función de su masa poblacional con el fin de determinar cuál de los tres métodos tiene un mejor ajuste al modelo planteado.

Con el fin de realizar los cálculos estadísticos se utilizará el entorno de desarrollo integrado (IDE) conocido como R Studio en dónde se podrán codificar las operaciones en el lenguaje de programación R.

Para poder medir la influencia que tienen los mencionados métodos estadísticos sobre el PBI se realizará una evaluación de cada Error porcentual absoluto medio (MAPE) generado.

Palabras clave: regresión, Producto Bruto Interno, redes neuronales, masa poblacional, error porcentual absoluto medio.

ABSTRACT

This thesis seeks, through the application of three statistical methods: linear regression, polynomial regression, and neural networks, to evaluate the influence that these have when developing a model that allows the forecast of the Gross Domestic Product (GDP) of Peru based on its population mass in order to determine which of the three methods has a better fit to the proposed model.

In order to carry out the statistical calculations, the integrated development environment (IDE) known as R Studio will be used, where the operations can be coded in the R programming language.

In order to measure the influence that the aforementioned statistical methods have on GDP, an evaluation of each mean absolute percentage error (MAPE) generated will be carried out.

Keywords: regression, Gross Domestic Product, neural networks, population mass, mean absolute percentage error.

CAPÍTULO I: INTRODUCCIÓN

1.1. Introducción

Resulta necesario iniciar mencionando que, al contener una amplia gama de factores constitutivos que representan distintas áreas de la actividad económica de un país, el Producto Bruto Interno (PBI) se convierte en una de las variables macroeconómicas más importantes para medir la capacidad de poder económico de un país. Ahora bien, teniendo en cuenta que el PBI es el valor acumulado de los bienes y servicios finales producidos por una economía, entiéndase país, y que es la población quien compra bienes y servicios finales (referido al cálculo mediante el método del gasto) cabe la necesidad de conocer de qué manera afecta la masa poblacional de un país al PBI de este.

Con relación a lo antes mencionado, resulta importante tener en cuenta que en los últimos años se han suscitado considerable cantidad de procesos migratorios a nivel mundial los cuáles han modificado la masa poblacional de algunos países, por su puesto el Perú es uno de ellos en donde un caso puntual y de mayor impacto es la migración de ciudadanos venezolanos. Para setiembre de 2022, 5.96 millones de venezolanos se han convertido en refugiados en América Latina y el Caribe, y 7.1 millones a nivel mundial (R4V Plataforma de Coordinación interagencial para Refugiados y Migrantes de Venezuela, 2022). Para agosto de 2021 el número de extranjeros que residían en el país ascendía a la cantidad 1 millón 3447 mil 893, siendo los más altos porcentajes de ingreso de extranjeros el 32.3% que ingresó durante el año 2018 y el 34.7% en el 2021. Así también, más de 86.8% de los extranjeros son de nacionalidad venezolana (Instituto

Nacional de Estadística e Informática, 2022). Así también, otro hecho importante que ha impactado en la masa poblacional mundial fue la pandemia del COVID-19 la cual elevó el número de defunciones. En el año 2020, por política estatal y recomendación de la Organización Mundial de la Salud, el Perú cerró sus fronteras debido a la pandemia de COVID 19 lo que detuvo el flujo de entrada y salida del país. Cabe mencionar también que, entre 2020 y 2021 los decesos por COVID 19 se estimaron el 14.9 millones a nivel mundial y en el Perú la cifra fue de más de 200 mil. (Ministerio de Salud, 2024)

1.2. 5

1.2.1. Determinación del problema

Resulta necesario conocer cuál de los tres métodos mencionados a continuación se ajusta mejor al pronóstico del PBI con respecto a la población. Primeramente, la regresión lineal predice una variable de respuesta cuantitativa a partir de una variable predictora cuantitativa, en segundo lugar, la regresión polinómica funciona de manera similar con la diferencia de que la relación se modela como una función polinomial de orden n, por último, las redes neuronales permiten establecer relaciones no lineales entre las variables de entrada y salida y tiene como principal potencial la detección de no-linealidades en series temporales. Por lo tanto, si bien los métodos antes mencionados son utilizados para pronosticar un valor, es necesario conocer cuál de ellos se ajusta de una manera más precisa en la predicción del valor.

En tal sentido, resulta problemático no conocer qué método estadístico se ajusta de manera más precisa a la relación entre la variación poblacional con la variación del PBI teniendo en cuenta la importancia que tiene este indicador, puesto que los gobernantes no tendrían una herramienta clave para la toma de decisiones en materia económica, además de la necesidad de conocer qué tanto influye la masa poblacional en el pronóstico del PBI.

1.2.2. Formulación del problema

1.2.2.1. Problema Principal

¿Cómo influyen la regresión simple, polinómica y redes neuronales en la calidad del modelo y pronóstico del PBI en función de la población?

1.2.2.2. Problemas secundarios

1. ¿Cómo influye un modelo basado en la población en la predicción del PBI?
2. ¿Cómo influyen la regresión simple en la calidad del pronóstico del PBI en función de la población?
3. ¿Cómo influye la regresión polinómica en la calidad del pronóstico del PBI en función de la población?
4. ¿Cómo influye las redes neuronales en la calidad del pronóstico del PBI en función de la población?

1.3. Objetivos

1.3.1. Objetivo principal

Determinar cómo influyen la regresión simple, polinómica y redes neuronales en la calidad del modelo y pronóstico del PBI en función de la población.

1.3.2. Objetivos secundarios

1. Determinar cómo influye un modelo basado en la población en la predicción del PBI
2. Determinar cómo influyen la regresión simple en la calidad del pronóstico del PBI en función de la población
3. Determinar cómo influye la regresión polinómica en la calidad del pronóstico del PBI en función de la población
4. Determinar cómo influye las redes neuronales en la calidad del pronóstico del PBI en función de la población

1.4. Importancia y alcance de la investigación

Resulta importante conocer en qué manera varía el PBI en función a la masa poblacional, dado que su pronóstico ayudaría, por ejemplo, en el planeamiento de ingresos y gastos del próximo año de un gobierno; el manejo más eficiente de la política monetaria al fijar de manera óptima la tasa de intervención en el mercado de acuerdo con el clima económico; la planeación de una empresa o institución que requiere conocer el valor futuro de la producción, entre otros.

Dado que las proyecciones en la tasa de crecimiento poblacional del Perú tienen una tendencia decreciente (Instituto Nacional de Estadística e Informática, 2024), determinar si la regresión simple, regresión múltiple o las redes neuronales se ajusta mejor el pronóstico del PBI a partir de la población permite no solo conocer una forma de estimar el mismo sino todas las ventajas que conlleva su predicción como las decisiones macroeconómicas a nivel de la administración pública y de futuras inversiones por parte de empresas, tanto nacionales como transnacionales, sino que también permite iniciar nuevas líneas de investigación para la estimación del PBI no solo como serie de tiempo sino como una variable dependiente de la población.

1.5. Limitaciones de la investigación

Es necesario mencionar que, si bien el PBI está influido por diversas variables, esta investigación se enfoca a la variable poblacional. Así también se tiene como limitante los datos históricos de la masa poblacional del Perú en un rango que va desde el año 1960 hasta el 2019.

CAPÍTULO II: REVISIÓN DE LITERATURA

2.1. Marco teórico

En esta sección se desarrollan cuatro puntos temas relevantes para la presente investigación, tales como el indicador macroeconómico, el PBI; los métodos estadísticos de pronóstico como el análisis de regresión lineal, el análisis de regresión polinomial y, de manera más profunda, las redes neuronales artificiales con el fin de poder facilitar la comprensión del objetivo principal de esta investigación: Determinar cómo influyen la regresión simple, polinómica y redes neuronales en la calidad del modelo y pronóstico del PBI en función de la población.

2.2. Antecedentes del estudio

Angulo (2013) en su investigación “Predicción del PBI real nacional trimestral: Redes neuronales autorregresivas vs. Metodología Box-Jenkins” abordó como problema principal el determinar cuál de las dos metodologías proporciona el mejor modelo de series de tiempo para el pronóstico del PBI real nacional trimestral, la hipótesis principal fue que el modelo de redes neuronales autorregresivas proporciona el mejor modelo para el pronóstico del PBI real trimestral de nuestro país, la población fue el PBI real nacional trimestral del Perú (en miles de nuevos soles del año 94) y tomó como muestra 74 datos de PBI trimestral que comprenden desde el I trimestre del año 1994 hasta el II trimestre del año 2012, como técnica de recolección de datos utilizó los datos publicados en la página web del Instituto Nacional de Estadística e Informática. La conclusión a la que llegó fue que, según las comparaciones realizadas entre los

modelos resultantes con ambas metodologías, se encontró que la metodología Box-Jenkins proporcionó el mejor modelo para el pronóstico PBI trimestral del Perú, tanto en el periodo de calibración (entrenamiento), en el de validación (evaluación), como en la serie completa.

Castro (2019) en su investigación “Modelo de pronóstico del comportamiento del índice del PBI en el Perú. enero 2007 – junio 2017” abordó como problema principal determinar cuál es el mejor modelo de pronóstico que describe el comportamiento del índice del PBI en el Perú, desde enero 2007 – junio 2017, utilizando la Metodología de Box y Jenkins, tuvo como objetivo principal construir un modelo de pronóstico que describa el comportamiento del Índice del PBI del Perú. 2007 – junio 2017, utilizando la metodología de BOX y JENKINS, la población fue la serie histórica del índice de PBI nacional en Perú, la muestra fue la serie histórica del índice de PBI en el Perú durante el periodo de enero 2007 hasta junio 2017, como técnica de recolección de datos utilizó los datos publicados en la página web del Instituto Nacional de Estadística e Informática. La conclusión a la que se llegó fue que el modelo que mejor se ajusta es SARIMA (3,1,1) (0,1,1) el cual presenta un MAPE de 0.5344099.

Espinoza (2022) en su investigación “Diseño de un sensor virtual para la predicción de la eficiencia de conversión de SO_2 a SO_3 en un reactor catalítico multi-lechos mediante redes neuronales artificiales” abordó como problema principal la posibilidad de predecir la eficiencia de conversión de una planta de ácido sulfúrico utilizando un sensor virtual basado en RNA, el objetivo principal

fue Diseñar un sensor virtual basado en RNAs para predecir la eficiencia de conversión de un reactor catalítico de una planta de ácido sulfúrico , la hipótesis fue que diseñar un sensor virtual basado en RNAs para predecir la eficiencia de conversión de un reactor catalítico de una planta de ácido sulfúrico, la población fueron los datos recolectados por el software ActiveFactory, la muestra fue los datos recolectados por el mencionado software en un periodo de tiempo de doce días, utilizó como método de recolección de datos utilizó la base de datos generada por el software ActiveFactory. La conclusión a la que se llegó fue que el sensor virtual basado en RNA (de estructura 9,9,1) tiene un mejor desempeño de predicción de la eficiencia de conversión en comparación con el sensor virtual derivado de leyes en estado estacionario.

Chávez (2022) en su investigación “Aplicación de redes neuronales artificiales para la valorización de los departamentos habitacionales de la ciudad de lima en 2020” abordó como problema principal el determinar cómo se avaluarán departamentos habitacionales en la ciudad de Lima aplicando un modelo de Redes Neuronales Artificiales, el objetivo principal fue Desarrollar un modelo de Redes Neuronales Artificiales para avaluar departamentos habitacionales, que permita mejorar el proceso de valorización de un departamento, tuvo como hipótesis que a través de la aplicación de un modelo de Redes Neuronales Artificiales se permitirá avaluar departamentos habitacionales en la ciudad de Lima, la población fue la data por el Banco Central de Reserva del Perú sobre Valorización de Departamentos agrupados por Trimestres (Primer Trimestre 1998 – Cuarto Trimestre 2020), la técnica de recolección de datos fue la

descarga de los datos publicados por la Gerencia Central de Estudios Económicos del Banco Central de Reserva del Perú (BCRP). La conclusión a la que se llegó fue que se demostró que los modelos de redes neuronales artificiales propuestos aprendieron de forma correcta haciendo una comparación entre los valores reales con los valores predichos por el modelo de red neuronal artificial. Esta comparación sirvió para obtener el valor de R, el cual fue 0.91743., este valor permite demostrar la gran capacidad de los modelos de redes neuronales artificiales para avaluar departamentos habitacionales en la ciudad de Lima en base a los datos brindados por el BCRP.

Quispe (2022) en su investigación “Implementación de redes neuronales artificiales para regresión en predicción de volumen de pedidos de textiles de moda” tuvo como objetivo principal el elaborar el protocolo de importaciones textiles para realizar el pedido del volumen óptimo en cantidad de metros de tela por container que optimice las ganancias, la población fue el registro de volúmenes de pedido de la empresa, la muestra fueron los volúmenes de pedidos de la empresa en el periodo 2016 – 2017, el método de recolección de datos fue la descarga de la base de datos de pedidos de la empresa. La conclusión a la que se llegó fue que se aprecia la capacidad del desempeño de las redes neuronales para modelar el sistema complejo de las variables que afectan en el éxito esperado de las ventas de los lotes importados según los menores errores obtenidos a comparación con modelos tradicionales.

2.3. Bases teóricas

2.3.1. Producto Bruto Interno (PBI)

Conocido también como el producto interno bruto (PIB) es la suma del valor agregado por empresas residentes, hogares y el gobierno que operan en una economía. (Peng, 2018).

Dicho de otra manera, el PBI suma distintos tipos de productos en una sola medida del valor de la actividad económica. Para tal fin utiliza lo precios del mercado, debido a que estos miden la cantidad que las personas pagarían por diferentes bienes (Mankiw, 2017)

El PIB mide el valor de la producción dentro de los límites geográficos de un país, esta producción tiene lugar en un periodo de tiempo, por lo general es un año o trimestre. Incluye todos los artículos producidos y vendidos de manera legal en los mercados. El PIB excluye a los artículos producidos y vendidos de forma ilícita como las drogas ilegales. Incluye bienes tangibles como alimentos, ropa o automóviles y también bienes intangibles como cortes de cabello, limpieza doméstica o consultas médicas. (Mankiw, 2017)

Se podría decir entonces que el PBI es una medida de la producción económica de un país. Se calcula sumando el valor de todos los bienes y servicios finales producidos en un territorio durante un período determinado, generalmente un año. El PBI es un indicador importante de la salud y la prosperidad de una economía

Para cuantificar el PBI existen tres métodos: Producción, Gasto e Ingreso. La producción se refiere a bienes y servicios; Gasto se refiere a consumo, inversión y exportaciones e Ingreso se refiere a remuneraciones, impuestos netos, consumos de capital fijo y excedentes neto de explotación. (Instituto Nacional de Estadística e Informática, 2000) Se resume lo antes mencionado en la Tabla 1. Métodos de cálculo del PBI.

Tabla 1 Métodos de cálculo del PBI

Método	**Descripción**	**Fórmula**
De la Producción	Producción de bienes y servicios	$PBI = \sum_{i=1}^{n} VAB + DM + Ip$
Del Gasto	Consumo, inversión y exportaciones	$PBI = GCH + GCG + FBKF + VE + X - M$
Del Ingreso	Remuneraciones, impuestos netos, consumo de capital fijo y excedente neto de explotación	$PBI = R + CKF + Ipm + EE$

Fuente: Elaboración propia basado en la publicación “Metodología de Cálculo del Producto Bruto Interno Anual” (Instituto Nacional de Estadística e Informática, 2000)

Donde:

VAB: $Valor\ Agregado\ Bruto$

DM: $Derechos\ de\ importación$

Ip: $Impuesto\ a\ los\ productos$

GCH: $Gasto\ de\ Consumo\ a\ los\ Hogares$

GCG: $Gasto\ de\ Consumo\ de\ Gobierno$

$FBKF$: $Formación\ Bruta\ de\ Capital\ Fijo$

VE: $Variación\ de\ Existencias$

X: $Exportaciones$

M: $Importaciones$

R: $Remuneraciones$

CKF: $Consumo\ de\ Capital\ Fijo$

Ipm: $Impuesto\ a\ la\ Producción\ e\ Importaciones$

EE: $Excedente\ de\ Explotación$

Algunas de las razones de la importancia del PBI incluyen:

1. Indicador de la economía: El PBI es una medida de la producción económica y del bienestar de un país, y es una de las formas más importantes de evaluar su desempeño económico.

2. Toma de decisiones gubernamentales: El PBI se utiliza como base para tomar decisiones gubernamentales sobre el gasto y la política fiscal.

3. Comparaciones internacionales: El PBI permite comparar el tamaño y el crecimiento de las economías de diferentes países e identificar tendencias y desafíos globales.

Evaluación de la calidad de vida: Aunque no mide directamente la calidad de vida de una sociedad, el PBI puede ser un indicador útil de la prosperidad económica y la capacidad de un país para proporcionar bienestar a sus ciudadanos.

PBI Nominal

Es el valor de la producción de bienes y servicios finales de un determinado territorio, a los precios de ese periodo. (Villamizar Jaimes, 2021) En periodos anuales el PBI nominal mide la producción utilizando los precios de cada año, también llamados precios corrientes, por lo tanto, su tasa de crecimiento puede ser resultado de una variación en los precios, en la cantidad producida o en ambas variables.

PBI Real

Es el valor de la producción de bienes y servicios finales de un determinado territorio, valorada a los precios de un año en específico denominado año base. (Cristancho Giraldo, 2023) También llamado PBI a precios constantes, la variación de su tasa se debe únicamente a las variaciones en la producción.

2.3.2. **Análisis de regresión**

El análisis de regresión en una de las técnicas más utilizadas para evaluar datos con varios factores. Su influencia y utilidad resultan del proceso lógico de usar una ecuación matemática para explicar la relación entre la variable de interés y un conjunto de variables predictoras. (Montgomery, Peck, & Vining, 2021) Se puede decir entonces que el análisis de regresión evalúa el comportamiento de una variable con el fin de establecer una relación y predecir el resultado de otra.

El análisis de regresión es utilizado en cualquier área de investigación en la que se esté interesado en explicar la relación entre una acción y su posible respuesta. (Gujarati, 2018) Utilizado en diversos campos como economía, finanzas, contabilidad, agricultura y demás el análisis de regresión es la base para estudiar la relación entre una variable de interés, denominada variable de respuesta, y un conjunto de variables predictoras.

2.3.2.1. **Regresión lineal simple:**

La regresión lineal simple establece la relación entre una variable dependiente y una variable explicativa. (Perez, Caso, Río, & López, 2011) La regresión lineal simple permite crear modelos en los que una única variable explica el comportamiento de otra. Mediante la técnica de regresión de una variable Y sobre una variable X se busca una función f(x) que se aproxime a una nube de puntos (X_i, Y_i). El modelo de regresión lineal simple tiene la siguiente expresión:

$$Y = \beta o + \beta X$$

En donde, βo es la ordenada en el origen es decir el valor que toma Y cuando X vale cero, β es la pendiente de la recta e indica la variación de Y al cambiar X en una unidad

2.3.2.2. **Regresión lineal múltiple:**

La regresión lineal múltiple establece la relación entre un conjunto de variables causales y una variable de respuesta. (Perez, Caso, Río, & López, 2011) En la regresión lineal múltiple se busca explicar la influencia de múltiples variables en una única variable de respuesta. Una cuestión de interés en la regresión lineal múltiple es saber del conjunto de variables predictoras cuáles de ellas son las que más influyen en el resultado.

El modelo de regresión lineal múltiple se puede expresar de la siguiente manera:

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \varepsilon$$

En donde, β_i son los coeficientes estimados del efecto marginal de cada X_i sobre Y.

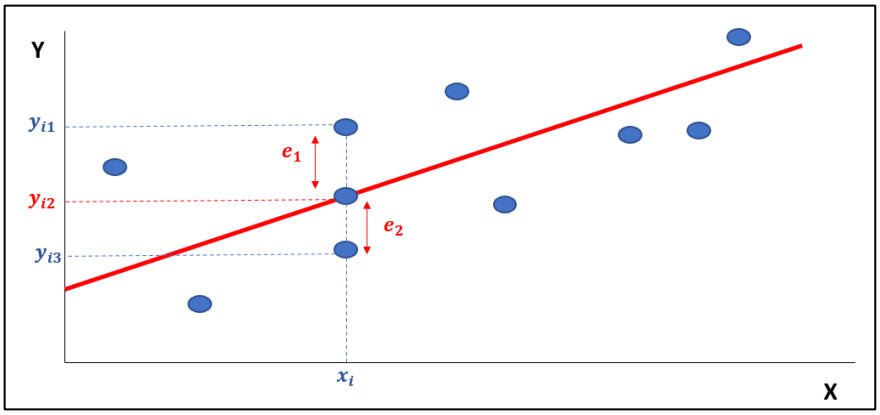
2.2.3.3. Recta de regresión mínimo-cuadrática

La diferencia entre el valor resultante de la recta Y_R y el valor teórico Y_i se denomina *error* o *residuo* el cual se denotará de la siguiente manera:

$$e = Y_i - Y_R$$

Gráficamente lo podemos observar en la Figura 1 en donde se puede apreciar que tal error puede ser negativo, positivo o nulo.

Figura 1 Representación gráfica del error



Fuente: Elaboración propia basado en el libro Aprendizaje automático y profundo en Python: Una mirada hacia la inteligencia artificial. (Pineda Pertuz, 2022)

Lo que se busca con la recta predictora es que la suma de estos errores sea mínima, existen ocasiones en las que este error tiene un valor negativo por lo que se podría anular con otro positivo proporcionando una idea falsa de ajuste. Para tal situación se utiliza el *ajuste por mínimos cuadrados* que consiste en minimizar la suma de los cuadrados de los errores.

Al elevar al cuadrado la expresión se elimina el signo y se penaliza a aquellos que tienen mayor cuantía. Para el caso de una regresión lineal simple

$$Y_R = \beta_0 + \beta X_i$$

La sumatoria de los cuadrados del error se puede expresar de la siguiente manera:

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta X_i)^2 = E(\beta_0, \beta)$$

Dado que el error se puede considerar una función E con los parámetros β_0 y β, y lo que se busca es minimizar dicha función para lo cual se necesita que la derivada de la función respecto a los parámetros sea nula, porque se igualarán a cero las derivadas parciales de β_0 y β, obteniendo así las siguientes ecuaciones:

$$\begin{cases} \dfrac{\delta E(\beta_0, \beta)}{\delta \beta_0} = -2 \displaystyle\sum_{i=1}^{n} (Y_i - \beta_0 - \beta X_i) = 0 \\ \dfrac{\delta E(\beta_0, \beta)}{\delta \beta} = -2 \displaystyle\sum_{i=1}^{n} (Y_i - \beta_0 - \beta X_i) X_i = 0 \end{cases}$$

De las derivadas parciales de los parámetros se desprenden el siguiente sistema de ecuaciones:

$$\begin{cases} \dfrac{\delta E(\beta_0, \beta)}{\delta \beta_0} \Rightarrow n\beta_0 + \beta \displaystyle\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \\ \dfrac{\delta E(\beta_0, \beta)}{\delta \beta} \Rightarrow \beta_0 \displaystyle\sum_{i=1}^{n} X_i + \beta \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i \end{cases}$$

Con el fin de obtener la solución a los parámetros β_0 y β el sistema de ecuaciones se expresa de manera matricial a fin de encontrar las soluciones mediante el método de Cramer:

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix}$$

Mediante el método de Cramer se pasa a despejar β

$$\beta = \frac{\Delta_\beta}{\Delta} = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i Y_i}{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}$$

Para el caso de β_0, este se expresará en función de β. A la ecuación $n\beta_0 + \beta\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$ se le dividirá por "n" teniendo en cuenta que

$$\frac{\sum_{i=1}^{n} X_i}{n} = Promedio\ de\ X = \bar{X}$$

Y que

$$\frac{\sum_{i=1}^{n} Y_i}{n} = Promedio\ de\ Y = \bar{Y}$$

La expresión quedaría de la siguiente manera:

$$\beta_0 = \bar{Y} - \beta\bar{X}$$

De esa manera se obtienen los coeficientes β_0 y β en la recta

$$Y_R = \beta_0 + \beta X_i$$

2.3.3. **Regresión polinomial**

Cuando se habla de regresión no lineal se refieren a modelos que no se pueden a justar mediante el método de los mínimos cuadrados. (Gernome, 1815)

La regresión polinomial permite predecir una variable respuesta a partir de una variable predictora, en dónde la relación se modela como una función polinomial de orden n. (Vinuesa, 2016) Pudiéndose representar de la siguiente manera:

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X^i + \varepsilon$$

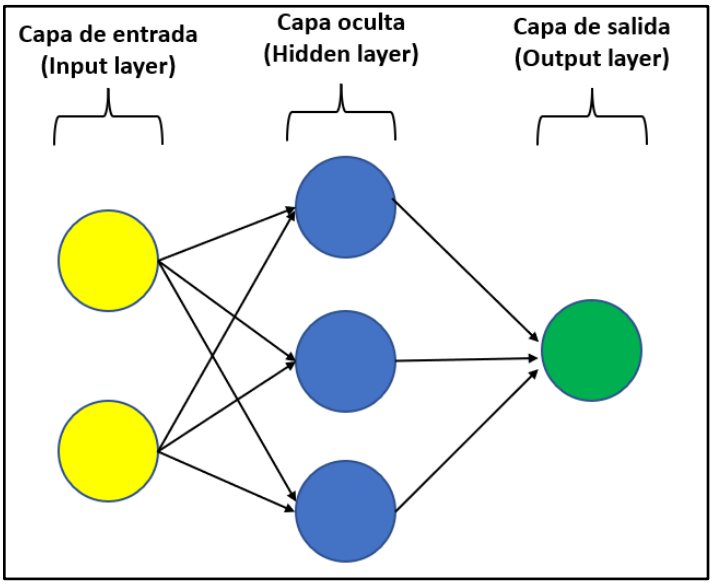
Se puede decir entonces que teniendo una sola variable explicativa X podemos ajustar la relación con potencias mayores como $X^2, X^3, \ldots, X^n$ de manera que describa mejor la curvatura de la relación Y, X.

2.3.4. **Red neuronal artificial**

2.3.4.1. Vista general

La red neuronal artificial (RNA) o en inglés *Artificial Neural Network* (ANN) es un sistema compuesto por varios elementos simples, pero altamente interconectados, en donde cada uno procesa información recibida (input) y genera una respuesta (output). (Cole, 2018) La estructura más simple de una red neuronal artificial está compuesta por un mínimo de 3 capas o *layers* estas son la capa de entrada o *input layer*, la capa oculta o *hidden layer* y la capa de salida o *output layer*. La red neuronal artificial se vuelve más compleja o profunda a medida que aumenta la cantidad de capas ocultas. En la Figura 2 se puede apreciar una red neuronal artificial en su forma más simple.

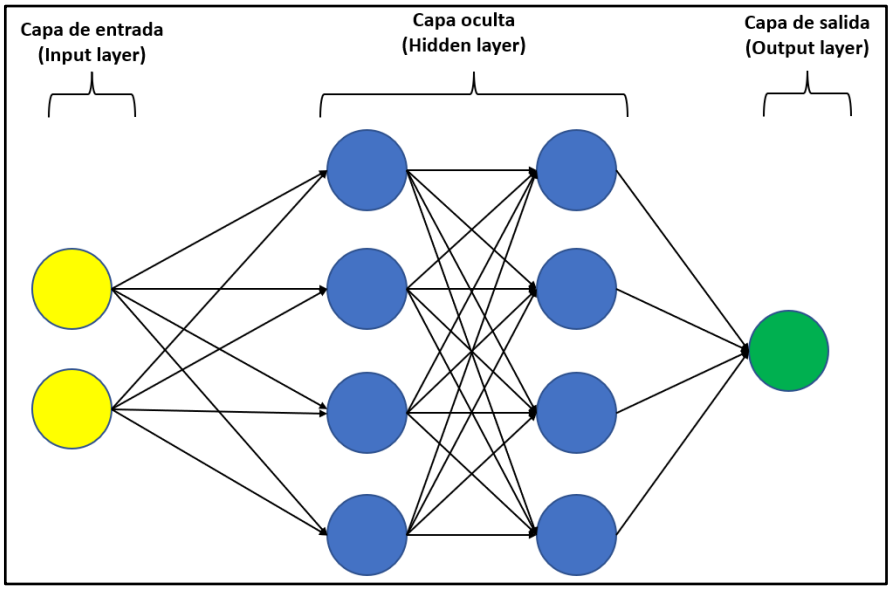
Figura 2 Red neuronal artificial simple



Fuente: Elaboración propia

En la Figura 3 podemos ver una red neuronal artificial más compleja o profunda.

Figura 3 Red neuronal compleja



Fuente: Elaboración propia

De manera homóloga a una red neuronal cerebral cada capa de la red neuronal artificial está compuesta por nodos o neuronas artificiales. Cada nodo o neurona artificial que conecta con otra tiene un peso y un valor umbral. (International Business Machines Corporation, 2020) Si la respuesta individual está por encima de un valor umbral específico entonces el nodo se activa y envía esa respuesta a la capa siguiente.

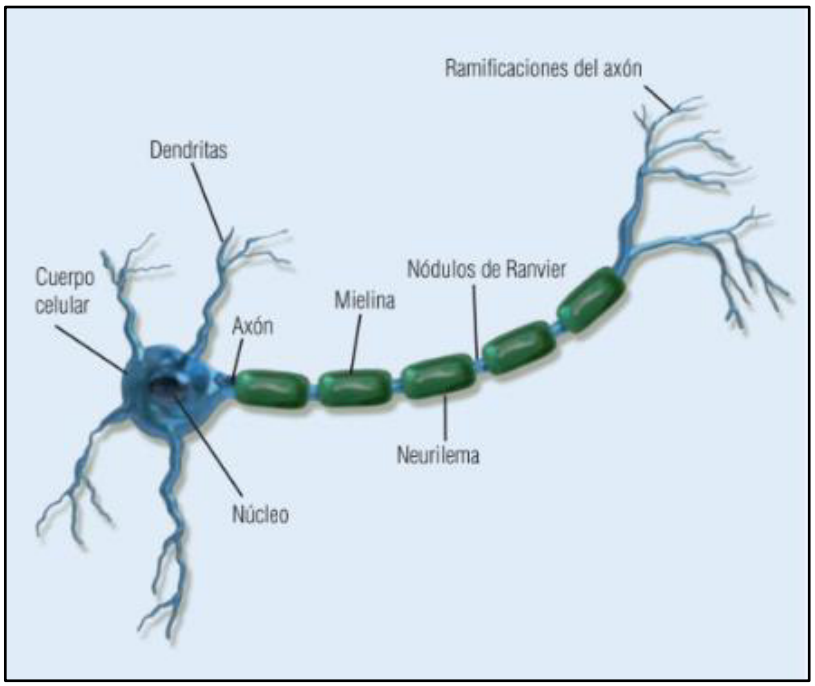
2.3.4.2. Fundamentos biológicos de las Redes Neuronales

El sistema de comunicación neuronal que recoge información del sistema nervioso y genera una respuesta se compone de receptores, el sistema nervioso y los órganos diana o efectores. (Isasi Viñuela & Galván León, 2004) Los receptores ubicados en las células sensoriales recogen información en forma de estímulos del ambiente o del interior del organismo, el sistema nervioso que

recibe esa información la procesa para enviarla a los órganos efectores, y los órganos diana o efectores como músculos y glándulas reciben la información procesada y la traducen en acciones hormonales, motoras, etc.

Un elemento estructural y esencial en el sistema de comunicación neuronal es la neurona. Considerada como la célula base del cerebro las neuronas ejecutan un proceso llamado sinapsis. (El-Sahili, 2022) Como se puede apreciar en la Figura 4 la neurona tiene distintas partes,

Figura 4 Partes de la neurona



Fuente: (Vásquez Conde, 2020, pág. 28)

La células nerviosas o neuronas son las más largas del cuerpo humano; sus axones pueden llegar a medir entre 0.1 y 1 m. El cuerpo celular dispone de protuberancias de longitud corta llamadas dendritas las cuales registran señales entrantes. Los axones transportan señales electroquímicas las cuales son

recibidas por las dendritas de otras neuronas en un proceso llamado sinapsis. (Müller-Esterl, 2008) Se puede decir entonces que las dendritas son las encargadas de recoger y propagar la información al interior del cuerpo celular quien la procesa hasta generar una respuesta que es canalizada por el axón en forma de una señal electroquímica. La señal propagada por el axón se ramifica y llega a las dendritas de otras neuronas en lo que se denomina sinapsis.

Se genera entonces una enorme malla de neuronas interconectadas en donde varía la concentración de iones en la sinapsis, esta concentración es muy importante dado que las neuronas no transportan una respuesta mediante sus axones solamente porque las dendritas recogieron la información. Las neuronas transmiten una repuesta de acuerdo con la saturación de iones, si esta sobrepasa un cierto umbral la respuesta es transportada caso contrario permanece inhibida.

2.3.4.3. Historia de la Red Neuronal Artificial (RNA)

Los primeros análisis acerca del comportamiento del cerebro y el proceso del pensamiento datan de la época de los filósofos griegos como Aristóteles y Platón, continuando en la Edad Media por con Descartes, sin embargo, fue en 1936 mediante el artículo *On computable numbers with an application to the entscheidungsproblem* que Alan Turing estudia el cerebro desde un punto de vista computacional. (Flores López & Fernández Fernández, 2008)

Las primeras investigaciones en Redes Neuronales Artificiales (RNA) datan de principios del siglo XIX pero no fue hasta la década de los 40 del siglo XX en

donde el estudio de RNA cobró fuerza la cual se ha ido incrementando hasta la actualidad. A continuación, se describirán brevemente los pasos más importantes de la historia de las RNA.

Warren McCulloch y Walter Pitts en 1943 realizaron el primer modelo matemático de una RNA en su artículo *A Logical Calculus of the Ideas Immanent in Nervous Activity*. El modelo McCulloch-Pitts opera con la premisa de que las neuronas operan con impulsos binarios y se introduce la idea de una función de paso por umbral la cual será usada en modelos posteriores.

Donald Hebb en su libro *Organiztion of Behavior* de 1949 desarrolla un paradigma de aprendizaje que ahora lleva su nombre – aprendizaje hebbiano-.

Marvin Misnky obtuvo los primeros resultados prácticos en RNA en 1951, junto a Edmons diseñaron la máquina *The stochastic Neural Anlog Reinforcement Calculator (SNARC)* con 40 neuronas, construida con tubos, motores y relés la máquina pudo modelar con éxito el comportamiento de una rata buscando comida en un laberinto.

En 1956, **Albert Uttley** creó una máquina teórica que ajustaba los parámetros de entrada utilizando la medida de entropía de Shanon. Posteriormente estas máquinas han sido utilizadas para la simulación de fenómenos atmosféricos.

En 1957, **Frank Rosenblatt** generalizó el modelo de células de McCulloch-Pitts añadiéndole aprendizaje, llamó a este modelo el **Perceptrón**. Logró desarrollar un modelo de dos niveles que ajustaba los pesos de las conexiones entre los niveles de entrada y salida en proporción al error entre la salida deseada y la salida obtenida.

Dos años después, **Bernard Widrow** diseñó una RNA similar al perceptrón, llamada *Adaptative Linear Element* o ADALINE el cual ha sido utilizado para el procesamiento adaptativo de señales.

Kunihiko Fukushima trabajó sobre RNA estudiando modelos espaciales y espacio-temporales para sistemas de visión y el cerebro. Notable por la creación de un paradigma de RNA multicapa para visión llamado COGNITRON.

Leon Cooper, ganador del premio Nobel, y **Charles Elbaum** formaron un grupo para el desarrollo, patente y explotación comercial de RNA llamado Nestor Associates en donde desarrollaron la red RCE (Reduced Coulomb Energy) con su propio sistema de aprendizaje NSL (Nestor Learning System)

En 1969, nuevamente **Marvin Minsky** junto con **Seymour Paper**, publicaron un análisis matemático del modelo de PERCEPTRON de Rosenblatt. Con este trabajo, demostraron las limitaciones del PERCEPTRON en la ejecución de funciones OR exclusivas o funciones XOR, también conocido como **el problema XOR de la separabilidad lineal**.

En 1974, **Paul Werbos** desarrolló el proceso de aprendizaje que hoy se conoce como **Backpropagation** o retroproagación. Hoy en día es de uso tan amplio que todos los softwares de RNa poseen módulos de backpropagation.

En 1986 **Terence Sejnowski** desarrolló el algoritmo de la máquina de Boltsman y su extensión de mayor orden, la primera RNA que reconocía un algoritmo de aprendizaje para red de tres niveles.

Por ese mismo año, **Robert Hecht-Nielsen** fue el principal diseñador de uno de los primeros computadores neuronales. Este neuro-computador el TRW MAR III.

2.3.4.4. Arquitectura de una RNA

El primer componente en la arquitectura de una RNA es la capa o *layer*, en la cual se encuentran las neuronas o nodos. (Perez & Gegúndez, 2021) Como se describió anteriormente existen 3 tipos de capas las de entrada, las ocultas y las de salida. La profundidad de una RNA es proporcional a la cantidad de capas ocultas que esta tiene.

Inicialización de pesos

Los algoritmos de entrenamiento necesitan un punto de partida desde el cual se modificarán los pesos de la red, el objetivo es que este punto se sitúe cerca del mínimo global y no se detenga en una zona de mínimos locales. (Perez & Gegúndez, 2021)

Es el proceso de asignar valores iniciales a los pesos en las conexiones entre las neuronas de la red. Los pesos determinan la importancia relativa de las entradas para cada neurona y, por lo tanto, son una parte importante de la forma en que la red procesa la información.

Hay diferentes métodos para inicializar los pesos, y el tipo de inicialización elegido puede afectar el rendimiento de la red y la capacidad de la red para aprender una representación útil de los datos.

Algunos de los métodos más comunes de inicialización de pesos incluyen:

- Inicialización aleatoria: los pesos se asignan a valores aleatorios dentro de un rango especificado, como una distribución normal o uniforme.

- Inicialización de Glorot: los pesos se asignan de manera que se garantice que la activación de las neuronas sea equilibrada tanto en la entrada como en la salida.

- Inicialización de He: una variante de la inicialización de Glorot que se utiliza en redes con funciones de activación rectificadoras.

La inicialización adecuada de los pesos puede ser clave para el éxito en el entrenamiento de la red y para evitar problemas como el gradients vanishing y exploding gradient.

Función de entrada

La función de entrada combina según corresponda las entradas x con sus pesos w. Las funciones de entradas más usadas son la sumatoria, producto y valor máximo de entrada. La función de entrada se puede representar matemáticamente como:

$$z = b + \sum(w * x)$$

Donde:

b es el sesgo o término de intercepto

w son los pesos asociados con los inputs

x son los valores de los inputs

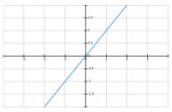
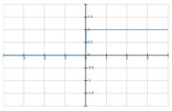
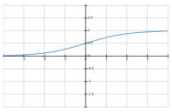
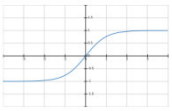
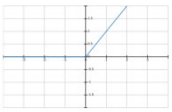
Σ es la operación de suma

z es la salida de la función de entrada

Función de Activación

Función matemática diseñada estratégicamente y seleccionada para mapear las entradas a una neurona direccionándolas hacia su salida. (Valbuena, 2021) En los comienzos, debido a sus aplicaciones limitadas se usaba la función escalonada, sin embargo, posteriormente se trabajó con valores reales, ya no 0 y 1. A pesar de que se podría utilizar cualquier función en la Figura 5 se muestran las funciones más utilizadas además de sus propiedades.

Figura 5 Principales funciones de activación

Nombre	Gráfica	Ecuación	Rango
Identidad		$f(x) = x$	$(-\infty, \infty)$
Escalonada o umbral		$f(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$	$\{0, 1\}$
Logística o sigmoide		$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$	$(0, 1)$
Tangente hiperbólica (tanh)		$f(x) = tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$	$(-1, 1)$
Unidad lineal rectificada (ReLU por sus siglas en inglés)		$f(x) = \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$	$[0, \infty)$

Fuente: (Pérez Borrero & Gegúndez Arias, 2021, pág. 44)

Tal como se puede ver en la Figura N. ° 5 existen distintas funciones de activación de uso común en las Redes Neuronales Artificiales:

- Identidad: se utiliza a menudo en problemas de regresión, donde el objetivo es predecir un valor numérico continuo en lugar de una clase. Al utilizar la función identidad de activación, se permite a la red neuronal aprender una relación lineal directa entre los inputs y la salida.

- Escalonada o Umbral: se utiliza a menudo en problemas de clasificación binaria, donde el objetivo es predecir si una entrada pertenece a una clase o a otra.

- Logística o Sigmoide: Es una función continua y suave que produce una salida que varía entre 0 y 1. Se utiliza a menudo en problemas de clasificación binaria.

- Tangente hiperbólica: Es similar a la sigmoide, pero produce una salida que varía entre -1 y 1. Se utiliza a menudo en problemas de clasificación multiclase.

- Rectificadora (ReLU): Es una función no lineal que produce una salida que es igual a la entrada si la entrada es positiva, y 0 si la entrada es negativa. Es una función de activación popular debido a su simplicidad y rapidez.

Se puede decir entonces que la función de activación es un componente clave de las redes neuronales, y su elección depende del problema a resolver y de los objetivos de la red neuronal. La función de activación se utiliza para modelar la respuesta de la neurona a la entrada y para determinar la activación o inactivación de la neurona

Valor de salida

Es el valor respuesta de la neurona que bien puede ser el valor final del modelo o el valor de entrada a una neurona de la siguiente capa.

El valor de salida de una red neuronal es el resultado final generado por la red una vez que se ha procesado la entrada. Este valor puede ser un número, un

vector o una distribución, dependiendo del problema que se esté abordando y de la arquitectura de la red neuronal.

En problemas de regresión, el valor de salida de una red neuronal es un número o un vector de números que representan la predicción de la red sobre el valor de una o más variables continuas.

En problemas de clasificación, el valor de salida de una red neuronal es un vector de probabilidades que representan la probabilidad de pertenencia a cada clase.

2.3.4.5. Aplicaciones de las RNA

Una de las ventajas que ofrecen las redes neuronales está en el uso de más una neurona en la capa de salida lo que permite abordar problemas por de clasificación multiclase. Por ejemplo, una RNA con 10 neuronas en la capa de salida puede ser capaz de identificar un dígito en una imagen. De manera más compleja, si se presenta una imagen como un vector de píxeles, esta puede ser procesada por una red neuronal pudiendo incluso ir más allá y generar una imagen como salida, lo único que se necesitaría es utilizar tantas neuronas de salida como píxeles se quiera que tenga la imagen.

En los últimos años se vienen publicando gran número de artículos científicos que tienen como base el uso de las redes neuronales y sus distintos tipos.

2.3.4.6. Paradigmas del aprendizaje de la RNA

Durante la fase de entrenamiento es posible, de formas distintas, ajustar los pesos de las RNA, en *Machine Learning* se suele utilizar algoritmos para este fin. (Perez & Gegúndez, 2021) En base al tipo de dato que se utilice se pueden abordar tres tipos distintos de paradigmas de aprendizaje:

Aprendizaje supervisado

Tipo de aprendizaje automático en el que se utiliza un conjunto de datos etiquetados para entrenar un modelo. En el aprendizaje supervisado, se conocen las entradas y las salidas deseadas para cada entrada, y el objetivo es construir un modelo que pueda generalizar y hacer predicciones precisas sobre nuevos datos.

En el aprendizaje supervisado, los datos se dividen en dos conjuntos: el conjunto de entrenamiento, que se utiliza para entrenar el modelo, y el conjunto de prueba, que se utiliza para evaluar la precisión del modelo.

Hay dos tipos principales de aprendizaje supervisado: la clasificación y la regresión. En la clasificación, el objetivo es predecir a qué clase pertenece una entrada. En la regresión, el objetivo es predecir una salida continua.

El aprendizaje supervisado es una técnica ampliamente utilizada en muchos problemas de inteligencia artificial, incluyendo la detección de spam, la identificación de objetos en imágenes, la predicción de precios de acciones y muchos otros.

Aprendizaje no supervisado

Tipo de aprendizaje automático en el que no se proporcionan etiquetas o salidas deseadas para los datos de entrada. En lugar de eso, el objetivo es encontrar patrones o relaciones en los datos sin una guía explícita.

En el aprendizaje no supervisado, el modelo aprende por sí solo a partir de los datos de entrada, sin la necesidad de un supervisor externo que proporcione información sobre la salida deseada.

Hay varios tipos de aprendizaje no supervisado, incluyendo clustering, reducción de dimensionalidad y aprendizaje profundo.

En el clustering, el objetivo es dividir los datos en grupos o clúster basados en similitudes. En la reducción de dimensionalidad, el objetivo es reducir la cantidad de características de los datos para facilitar su análisis. En el aprendizaje profundo, se utilizan redes neuronales para aprender características representativas a partir de los datos de entrada.

El aprendizaje no supervisado es una técnica ampliamente utilizada en muchos problemas de inteligencia artificial, incluyendo la compresión de datos, la segmentación de imágenes y la detección de anomalías.

Aprendizaje por refuerzo

Tipo de aprendizaje automático en el que un agente interactúa con su entorno para aprender a tomar acciones que maximizan una recompensa. En el aprendizaje por refuerzo, el agente toma acciones, observa las consecuencias y recibe una recompensa o un castigo en función de su desempeño. A continuación, utiliza esta información para mejorar su estrategia y tomar acciones más acertadas en el futuro.

El aprendizaje por refuerzo es una técnica ampliamente utilizada en problemas de control, juegos y robótica. Por ejemplo, se puede utilizar para entrenar un agente para controlar un vehículo autónomo o para jugar un juego de estrategia.

En el aprendizaje por refuerzo, el agente aprende a través de la interacción con su entorno y la retroalimentación de la recompensa. El objetivo es encontrar una política óptima que maximice la recompensa total obtenida por el agente a lo largo del tiempo.

El aprendizaje por refuerzo es una técnica avanzada y desafiante, pero también es una de las más potentes y versátiles en el campo de la inteligencia artificial.

2.3.4.7. Backpropagation

Backpropagation es un algoritmo de entrenamiento para redes neuronales artificiales. Se utiliza para ajustar los pesos de las conexiones entre las neuronas de una red neuronal con el objetivo de minimizar una función de pérdida o costo.

El algoritmo funciona de la siguiente manera:

1. Se proporciona una entrada a la red neuronal y se propaga a través de la red, generando una salida.

2. Se calcula la diferencia entre la salida real y la salida deseada, lo que representa el error de la red.

3. Se realiza una propagación hacia atrás, es decir, se calcula el gradiente de la función de pérdida con respecto a cada uno de los pesos de la red.

4. Se ajustan los pesos de la red en la dirección opuesta al gradiente, utilizando una técnica de optimización como el gradiente descendente estocástico.

5. Se repiten los pasos 1 a 4 para cada muestra del conjunto de entrenamiento.

Backpropagation es un algoritmo eficaz para el entrenamiento de redes neuronales y es ampliamente utilizado en muchos problemas de aprendizaje automático, incluyendo clasificación, regresión y procesamiento de lenguaje natural.

La idea de este algoritmo tiene como base la regla de la cadena, utilizada en el cálculo de la derivada de una composición de funciones, básicamente lo que es una red neuronal. (Perez & Gegúndez, 2021) De esta manera es posible calcular la gradiente en función de la pérdida, no solo para los pesos de las neuronas de la capa de salida sino para toda la red.

El algoritmo de *backpropagation* cuenta con dos etapas, la primera conocida como *forward pass*, en donde se obtiene el valor de pérdida junto con el resultado de todas las operaciones intermedias de la red. La segunda etapa es conocida como *backward pass*, al obtener el valor de perdida en el paso anterior se calcula la gradiente de la función de pérdida.

2.4. Definición de términos

Indicador

Medida que se utiliza para cuantificar la eficiencia y/o eficacia de una actividad o proceso. (Heredia Alvaro, 2001)

Actividad económica

Operaciones económicas en las que combinando mano de obra, equipos, materias primas e insumos con el fin de producir bienes y servicios. (Instituto Nacional de Estadística e Informática, 2013)

Población

Conjunto de seres humanos que viven en un espacio geográfico determinado en un tiempo determinado. (Malthus, 1846)

Modelo estadístico

Ecuación Matemática que codifica información extraída de los datos. (International Business Machines Corporation, 2021)

Sesgo

Desviación sistemática de la norma o racionalidad en la toma de decisiones. Tendencia a desviarse de la objetividad. (Kahneman, 2011)

Conjunto de datos (data set)

Grupo de registros almacenados en un tipo de archivo que contienen información sobre un determinado tema. (International Business Machines Corporation, 2016)

Entrenamiento del modelo

Se trata de contrastar los resultados del modelo, no solo con los datos usados para la estima sino también con los datos de validación. (Pajares Mantinsanz & De la Cruz García, 2022)

Algoritmo

Sucesión finita de instrucciones diseñadas para obtener un resultado. (Marquina Vila, 2017)

Inteligencia Artificial

Capacidad de una máquina de poder percibir, asociar, predecir, planificar y controlar con el fin de lograr un objetivo. (Boden, 2017)

Aprendizaje automático (machine learnig)

Secuencia de procesos enfocados a desarrollar un modelo exitoso de predicción.

(Priyadarshini Jeyaraman, Renbo Olsen, & Wambugu, 2019)

CAPÍTULO III: HIPÓTESIS Y VARIABLES

3.1. Hipótesis

3.1.1. Hipótesis principal

La regresión simple, polinómica y redes neuronales influyen favorablemente calidad del modelo y pronóstico del PBI en función de la población

3.1.2. Hipótesis secundarias

- Un modelo basado en la población explica en gran medida la predicción del PBI.
- La regresión simple influye favorablemente en la calidad del pronóstico del PBI en función de la población.
- La regresión polinómica influye favorablemente en la calidad del pronóstico del PBI en función de la población.
- Las redes neuronales influyen favorablemente calidad del pronóstico del PBI en función de la población.

3.2. Variables

En la Tabla 2 se presentan las variables y su relación de dependencia con cada hipótesis.

Tabla 2 Identificación de variables

Variable	Hipótesis	Relación de dependencia
Modelo basado en la población	Un modelo basado en la población explica en gran medida la predicción del PBI	Independiente

Variable	Hipótesis	Relación de dependencia
Modelo de regresión simple	La regresión simple influye favorablemente en la calidad del pronóstico del PBI en función de la población	Independiente
Regresión polinómica	La regresión polinómica influye favorablemente en la calidad del pronóstico del PBI en función de la población	Independiente
Red neuronal backpropagation	Las redes neuronales influyen favorablemente en la calidad del pronóstico del PBI en función de la población	Independiente
Calidad del pronóstico	Un modelo basado en la población explica en gran medida la predicción del PBI	Dependiente
Calidad del pronóstico	La regresión polinómica influye favorablemente en la calidad del pronóstico del PBI en función de la población	Dependiente
Calidad del pronóstico	Las redes neuronales influyen favorablemente en la calidad del pronóstico del PBI en función de la población	Dependiente

3.3. Operacionalización de las variables

En la Tabla 3 se presentan las variables y su definición operacional.

Tabla 3 Definición operacional

Variables	**Definición conceptual**	**Definición operacional**
Modelo basado en la población	El modelo con una variable independiente es uno de los modelos más básicos en el análisis de regresión, y se utiliza para establecer una relación lineal entre una variable dependiente y una variable independiente. **Fuente especificada no válida.**	Fases de implementación: Diseño de modelo y evaluación de resultados
Modelo de regresión simple	La regresión lineal simple establece la relación entre una variable dependiente y una variable explicativa. (Perez, Caso, Río, & López, 2011)	Fases de implementación: Diseño de modelo y evaluación de resultados
Modelo de regresión polinómica	La regresión polinómica predice una variable respuesta a partir de una variable predictora, en dónde la relación se modela como una función polinomial de orden n. (Vinuesa, 2016)	Fases de implementación: Diseño de modelo y evaluación de resultados
Modelo de red neuronal backpropagation	La red neuronal artificial (RNA) es un sistema compuesto por varios elementos simples, pero altamente interconectados, en donde cada uno procesa información recibida y genera una respuesta. (Cole, 2018)	Fases de implementación: Diseño de modelo y evaluación de resultados
Calidad del pronóstico	La capacidad de un modelo de pronóstico para producir estimaciones precisas y confiables de una variable futura. **Fuente especificada no válida.**	Evaluación de resultados

Fuente: Elaboración propia

Así también, los resultados de los pronósticos de cada modelo serán evaluados mediante el Error Porcentual Absoluto Medio (MAPE), el cual deberá situarse en un intervalo determinado. Lo antes mencionado se resumen en la Tabla 4.

Tabla 4 Análisis de los resultados

Variable	**Indicador**	**Escala**	**Estadístico de contraste**
Calidad de pronóstico	Error Porcentual Absoluto Medio (MAPE)	Intervalo	Error Porcentual Absoluto Medio (MAPE)

Así también, se presenta la Matriz de Consistencia en la Tabla 5.

Tabla 5 Matriz de consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLE INDEPENDIENTE (VI)	VARIABLE DEPENDIENTE (VD)	INDICADOR (VI)	INDICADOR (VD)
¿Cómo influyen la regresión simple, polinómica y redes neuronales en la calidad del modelo y pronóstico del PBI en función de la población?	Determinar cómo influyen la regresión simple, polinómica y redes neuronales en la calidad del modelo y pronóstico del PBI en función de la población	La regresión simple, polinómica y redes neuronales influyen favorablemente calidad del modelo y pronóstico del PBI en función de la población				
¿Cómo influye un modelo basado en la población en la predicción del PBI?	Determinar cómo influye un modelo basado en la población en la predicción del PBI	Un modelo basado en la población explica en gran medida la predicción del PBI	Modelo basado en la población	Calidad del pronóstico	Modelo basado en la población	Error Porcentual Absoluto Medio (MAPE)

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLE INDEPENDIENTE (VI)	VARIABLE DEPENDIENTE (VD)	INDICADOR (VI)	INDICADOR (VD)
¿Cómo influye la regresión simple en la calidad del pronóstico del PBI en función de la población?	Determinar cómo influyen la regresión simple en la calidad del pronóstico del PBI en función de la población	La regresión simple influye favorablemente en la calidad del pronóstico del PBI en función de la población	Modelo de regresión simple	Calidad del pronóstico	Modelo de regresión simple	Error Porcentual Absoluto Medio (MAPE)
¿Cómo influye la regresión polinómica en la calidad del pronóstico del PBI en función de la población?	Determinar cómo influye la regresión polinómica en la calidad del pronóstico del PBI en función de la población	La regresión polinómica influye favorablemente calidad del pronóstico del PBI en función de la población	Modelo de regresión polinómica	Calidad del pronóstico	Modelo de regresión polinómica	Error Porcentual Absoluto Medio (MAPE)

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLE INDEPENDIENTE (VI)	VARIABLE DEPENDIENTE (VD)	INDICADOR (VI)	INDICADOR (VD)
¿Cómo influyen las redes neuronales en la calidad del pronóstico del PBI en función de la población?	Determinar cómo influye las redes neuronales en la calidad del pronóstico del PBI en función de la población	Las redes neuronales influyen favorablemente calidad del pronóstico del PBI en función de la población	Modelo de red neuronal *backpropagation*	Calidad del pronóstico	Red neuronal *backpropagation*	Error Porcentual Absoluto Medio (MAPE)

CAPÍTULO IV: MATERIALES Y MÉTODOS

4.1. Área de estudio

El área es la masa poblacional y el PBI del Perú desde el año 1994 hasta el 2020. Además, las áreas del conocimiento de la Organización para la Cooperación y Desarrollo Económicos (OCDE), adoptadas por CONCYTEC, en las que se desarrolla el presente trabajo se detallan en la Tabla 6.

Tabla 6 Áreas del conocimiento abordadas

N. °	Área	Subárea	Disciplina	Código
1	Ciencias naturales	Matemáticas	Estadísticas y probabilidades	1.01.03
2	Ciencias sociales	Economía y negocios	Econometría	5.02.03
3	Ciencias sociales	Sociología	Demografía	5.04.02

4.2. Tipo y diseño de investigación

La presente investigación es del tipo aplicada y tiene un diseño descriptivo - explicativo donde a partir del establecimiento de la variable independiente se observa la influencia sobre la variable dependiente.

El esquema del diseño es el siguiente:

M – O – Mod

En donde:

M: Muestra

O: Observaciones

Mod: Modelo de pronóstico 1, 2 y 3

4.3. Población y muestra

Para esta investigación está conformada por todos los datos históricos referentes al PBI y la dimensión poblacional en el Perú; por lo tanto, es infinita.

La muestra de estudio en esta investigación está constituida por los datos históricos del PBI en el Perú, abarcando el período comprendido desde el año 1994 hasta el año 2020.

4.4. Procedimientos, técnicas e instrumentos de recolección de datos

Para llevar a cabo la recopilación de los registros correspondientes a los datos históricos del PBI en el Perú se utilizará la técnica de revisión de base de datos, la cual implica un minucioso proceso de búsqueda, selección y verificación de información relevante. En este sentido, se empleará la hoja de cálculo de Microsoft Excel, la cual tiene una capacidad para organizar y procesar datos de manera eficiente y precisa.

La elección de esta herramienta se debe a sus características versátiles y su amplio uso en el ámbito de la investigación económica. Por consiguiente, se procederá a descargar la información requerida desde el portal del Instituto Nacional de Estadística e Informática (INEI). Los registros obtenidos incluirán

información detallada sobre la población y el PIB durante el periodo de estudio, abarcando desde el inicio de la muestra en el año 1994 hasta el año 2020, asegurando así una amplia cobertura temporal para un análisis exhaustivo y riguroso.

4.5. Análisis estadístico

Las etapas del análisis estadístico son las siguientes:

Etapa 1. Obtención de datos:

La extracción de la base de datos se realiza desde el portal web de estadísticas del Instituto Nacional de Estadística e Informática. En las secciones de series anuales de PBI y Población.

Etapa 2. Preparación de la base de datos

Las bases de datos extraídas se descargan en formato *.xlsx*, formato perteneciente a la hoja de cálculo Microsoft Excel.

a) Limpieza de datos

   Búsqueda, corrección o limpieza de datos nulos u *outliers*.

b) Categorización de variables

   Numeración de registros y asignación de nombres a las columnas.

Etapa 3. Correr los modelos

a) En el entorno de trabajo R Studio, los datos son utilizados en la ejecución del modelo de regresión lineal simple.
b) En el entorno de trabajo R Studio, los datos son utilizados en la ejecución del modelo de regresión lineal polinómica.

c) En el entorno de trabajo R Studio, los datos son utilizados en la ejecución del modelo de Red Neuronal Artificial (Se define una estructura de Red Neuronal Artificial de 1 neurona en la capa de entrada, 3 neuronas en la capa oculta y una neurona en la capa de salida.

d) El script de cada modelo se presenta en el apartado **5.7 Script utilizado**.

Etapa 4. Valores de salida

Se obtiene el valor estimado del PBI en un año determinado.

CAPÍTULO V: RESULTADOS

5.1. Preparación de datos

Se muestra en la Tabla 7 la base de datos utilizada como muestra, la misma que muestra los años, el PBI (en millones de soles, valores a precios corrientes) y a la población (cantidad de habitantes)

Tabla 7 Base de datos

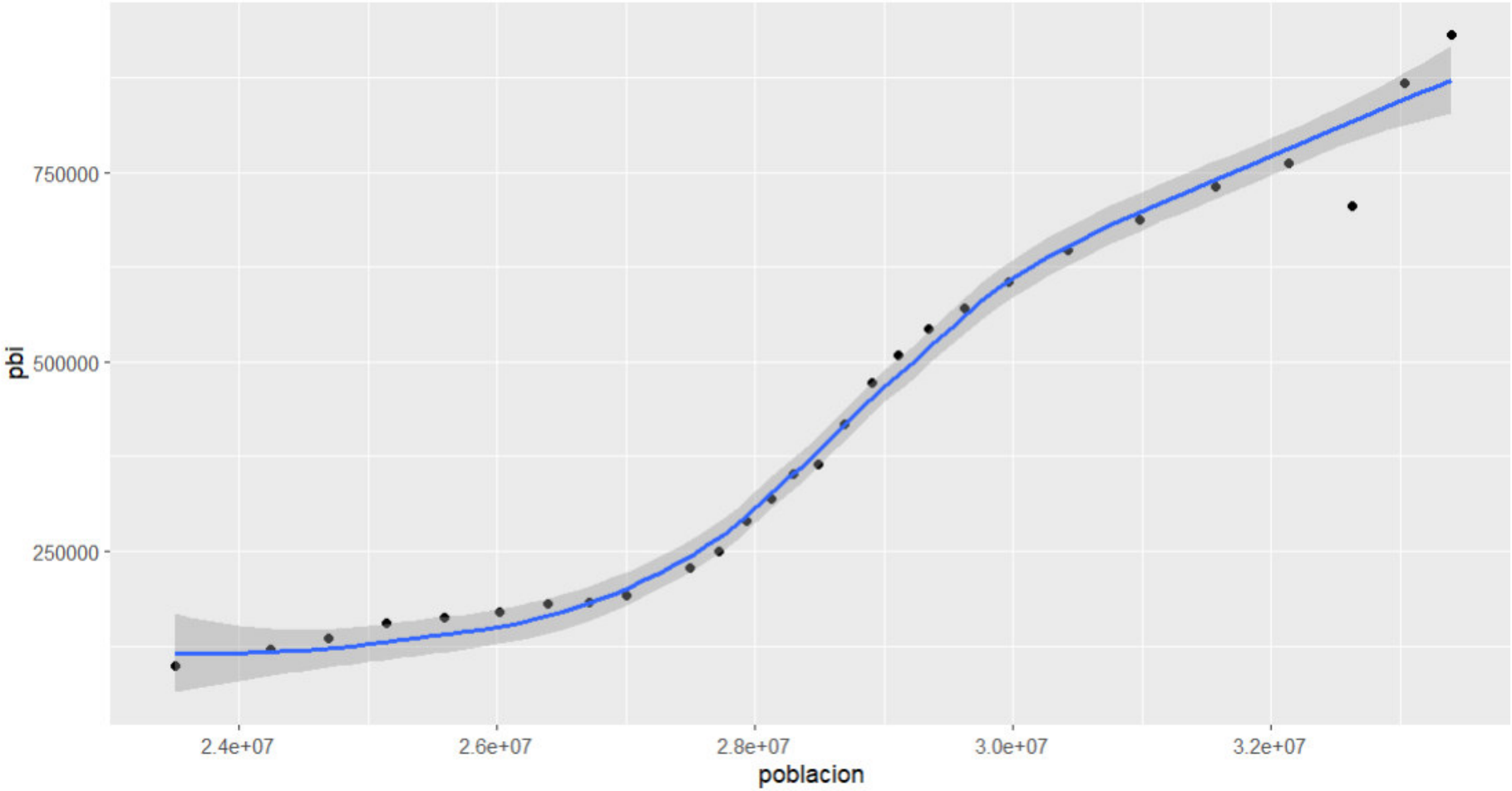
Año	PBI	Población
1994	98579	23501974
1995	120263	24242600
1996	135606	24689213
1997	154905	25145317
1998	162586	25592876
1999	169859	26013829
2000	180584	26390142
2001	182527	26714547
2002	192691	26999085
2002	192691	26999085
2004	227935	27492091
2005	250749	27722342
2006	290271	27934784
2007	319693	28122158
2008	352719	28300372
2009	363943	28485319
2010	416784	28692915
2011	473049	28905725
2012	508131	29113162
2013	543556	29341346
2014	570041	29616414
2015	604416	29964499

Año	PBI	Población
2016	647668	30422831
2017	687989	30973992
2018	731588	31562130
2019	761984	32131400
2020	705795	32625948
2021	868149	33035304
2022	930536	33396698

Fuente: (Instituto Nacional de Estadística e Informática, 2022)

Al graficar las columnas de PBI y Población de la Tabla anterior, se obtiene la Figura 6 en la cual se puede apreciar desde ya que un ajuste adecuado para estos valores no contemplaría a la regresión lineal, sino que más bien a la polinómica.

Figura 6 Gráfico PBI vs Población



Fuente: Elaboración propia basada en los datos de la Tabla 7 Base de datos

Ahora es necesario dividir la base de datos en dos partes: una para el entrenamiento de los modelos y la segunda para la prueba de estos, esta división se hará en porcentajes de 80% y 20%, respectivamente. Lo antes mencionado se muestra en la Tabla 8 y la Tabla 9, respectivamente.

Tabla 8 Base de datos de entrenamiento

Año	PBI	Población
1994	98579	23501974
1995	120263	24242600
1996	135606	24689213
1997	154905	25145317
1998	162586	25592876
1999	169859	26013829
2000	180584	26390142
2001	182527	26714547
2002	192691	26999085
2003	204337	27254632
2004	227935	27492091
2005	250749	27722342
2006	290271	27934784
2007	319693	28122158
2008	352719	28300372
2009	363943	28485319
2010	416784	28692915
2011	473049	28905725
2012	508131	29113162
2013	543556	29341346
2014	570041	29616414
2015	604416	29964499
2016	647668	30422831
2017	687989	30973992

Fuente: Elaboración propia basada en los datos de la Tabla 7 Base de datos

Tabla 9 Base de datos de prueba

Año	PBI	Población
2018	731588	31562130
2019	761984	32131400
2020	703915	32625948
2021	878380	33035304
2022	945329	33396698

Fuente: Elaboración propia basada en los datos de la Tabla 7 Base de datos

5.2. Modelo Lineal

El modelo lineal fue generado con los datos de la Tabla 8. Se presentan los resultados en la Tabla 10.

Tabla 10 Resultados del modelo lineal

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.06E+06	2.00E+05	-10.32	6.74E-10	***
Población	8.67E-02	7.23E-03	11.99	4.02E11	***

Signif. Codes: 0'***' 0.001 '**' 0.01'*' 0.05 '.' 0.1' '
Residual standard error: 68830 on 22 degrees of freedom
Multiple R-squared: 0.8673, Adjusted R-squared: 0.8613
F-statictic: 143.8 on 1 and 22 DF, p-value:4.019e-11

Fuente: Elaboración propia

Como siguiente paso se va a analizar la significancia tanto de β_0 como de β_1, para ello las hipótesis estadísticas en ambos casos son:

H_0: $\beta = 0$

H_1: $\beta \neq 0$

α: 0.05

Conforme lo observado en la columna Pr(>|t|) de la Tabla antes vista, en ambos casos se rechaza H_0, por lo tanto, todos los coeficientes tienen significancia en el modelo.

Así también, para evaluar la significancia global del modelo, se tienen las siguientes hipótesis:

H_0: El modelo no es estadísticamente significativo

H_1: El modelo es estadísticamente significativo

α: 0.05

Como se observa en la Tabla 10, el valor p global es de 4.019e-11, por lo tanto, se rechaza H_0, luego el modelo es estadísticamente significativo.

Finalmente se puede apreciar que el coeficiente R^2 ajustado (*Adjusted R-quared*) es de 0.8613, es decir, un 86% de la variabilidad de los resultados es explicado por el modelo, lo cual es un porcentaje alto y por lo tanto aceptable.

Luego de la revisión de los principales indicadores del modelo, se puede observar que la cantidad de población tiene un impacto relevante en el PBI, además el modelo parecería adecuado para el pronóstico, sin embargo, es necesario evaluar el cumplimiento de los supuestos de este tipo de modelos;

para ello se va a analizar la normalidad de los residuos, la homocedasticidad, la independencia de errores y la linealidad del modelo.

Con respecto a la prueba de normalidad de residuos, debido a la cantidad de estos se aplicó la prueba de Shapiro-Wilk, con las siguientes hipótesis estadísticas:

H_0: Los residuos siguen una distribución normal

H_1: Los residuos no siguen una distribución normal

α: 0.05

Los resultados fueron los siguientes: W = 0.92175, p-value = 0.06384, como se puede apreciar, el valor p no cumple con la regla de rechazo de H_0 por lo tanto, los residuos siguen una distribución normal, con lo cual cumplen con este supuesto.

El siguiente supuesto a comprobar es el de la homocedasticidad, para ello se aplicó la prueba de Breusch-Pagan, con las siguientes hipótesis estadísticas:

H_0: Existe homocedasticidad

H_1: No existe homocedasticidad

α: 0.05

Los resultados fueron los siguientes: BP = 1.6958, df = 1, p-value = 0.1928; similar al caso anterior, el valor p no cumple con la regla de rechazo de H_0 por lo tanto, existe homocedasticidad, con lo cual se cumple también con este supuesto.

A continuación, se va a verificar que no exista autocorrelación, es decir, se va a verificar la independencia de los errores, para ello se va a utilizar la prueba de Durbin Watson, con lo cual se tienen las siguientes hipótesis estadísticas:

H_0: No existe autocorrelación

H_1: Existe autocorrelación

α: 0.05

Los resultados obtenidos fueron:

lag Autocorrelation D-W Statistic p-value

1 0.8568681 0.09892911 0

Alternative hypothesis: rho != 0

Se verifica que el valor p cumple con la regla de rechazo de H_0 por lo tanto, existe autocorrelación.

Finalmente, queda verificar la linealidad del modelo, para ello se va a aplicar la prueba RESET, cuyos resultados fueron:

H_0: El modelo presenta linealidad

H_1: El modelo no presenta linealidad

α: 0.05

RESET = 35.633, df1 = 2, df2 = 20, p-value = 2.554e-07

En esta prueba el valor P cumple con la regla de rechazo de H_0, por lo tanto, se necesita un modelo de orden superior

5.3. Modelo Polinómico

En el apartado anterior se ha verificado la necesidad de generar un modelo de orden superior; esto debido a los resultados de la prueba Durbin – Watson y RESET, en ese sentido se procede a generar modelos de orden superior, los cuales se someterán a la prueba de los supuestos antes aplicados. En la Tabla 11 se presenta el resumen de los resultados de los modelos desde el segundo al sexto orden.

Tabla 11 Resultados de los modelos de orden 2 al 6

Orden del modelo polinómico		**2°**	**3°**	**4°**	**5°**	**6°**
Supuesto de normalidad de residuos (Prueba Shapiro Wilk)	**Valor de P**	0.8696	0.07818	0.08768	0.5421	0.4744
	Estado	Cumple	Cumple	Cumple	Cumple	Cumple
Supuesto de Homocedasticidad (Prueba Breusch – Pagan)	**Valor de P**	0.02411	0.72	0.6344	0.1803	0.6082
	Estado	No cumple	Cumple	Cumple	Cumple	Cumple
Supuesto de Independencia de errores (Prueba de Durbin – Watson)	**Valor de P**	0	0	0.002	0	0.002
	Estado	No cumple	No cumple	No cumple	No cumple	No cumple
Supuesto de linealidad del modelo (Prueba RESET)	**Valor de P**	5.43E-07	3.75E-08	5.65E-07	1.21E-06	0.01101
	Estado	No cumple	No cumple	No cumple	No cumple	No cumple

Fuente: Elaboración propia

Como se puede apreciar en la Tabla 11, los modelos polinómicos desde el grado 2 al 6 no cumplen con los supuestos establecidos anteriormente. Se presenta en la Tabla 12 el modelo polinómico de grado 7.

Tabla 12 Modelo polinómico de orden 7

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	327453	1551	211.156	< 2.00E-16	***
poly(Población,7)1	825460	7597	108.654	< 2.00E-16	***
poly(Población,7)2	280969	7597	36.983	< 2.00E-16	***
poly(Población,7)3	-49359	7597	-6.497	7.36E-06	***
poly(Población,7)4	-134244	7597	-17.67	6.4E-12	***
poly(Población,7)5	-20204	7597	-2.659	0.0171	*
poly(Población,7)6	55638	7597	7.324	1.71E-06	***
poly(Población,7)7	19738	7597	2.598	0.0194	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7597 on 16 degrees of freedom
Multiple R-squared: 0.9988, Adjusted R-squared: 0.9983
F-statistic: 1942 on 7 and 16 DF, p-value: < 2.2e-16

Fuente: Elaboración propia

En la Tabla 12 se puede observar que todos los coeficientes son significativos, también se aprecia que el modelo tiene un coeficiente R^2 ajustado de 99.83% y un valor p que lo cataloga como un modelo estadísticamente significativo. A continuación, en la Tabla 13 se resumen los resultados del modelo al aplicar las pruebas y comprobar los supuestos.

Tabla 13 Resumen de verificación de supuestos

Supuesto	**Prueba**	**Valor p**	**Estado**
Normalidad de residuos	Shapiro – Wilk	0.7189	Cumple
Homocedasticidad	Breusch – Pagan	0.986	Cumple
Independencia de errores	Durbin – Watson	0.116	Cumple

Supuesto	Prueba	Valor p	Estado
Linealidad del modelo	RESET	0.151	Cumple

Como se puede observar en la Tabla 13, el modelo polinómico de orden 7 cumple con todos los supuestos, a continuación, se presenta su ecuación:

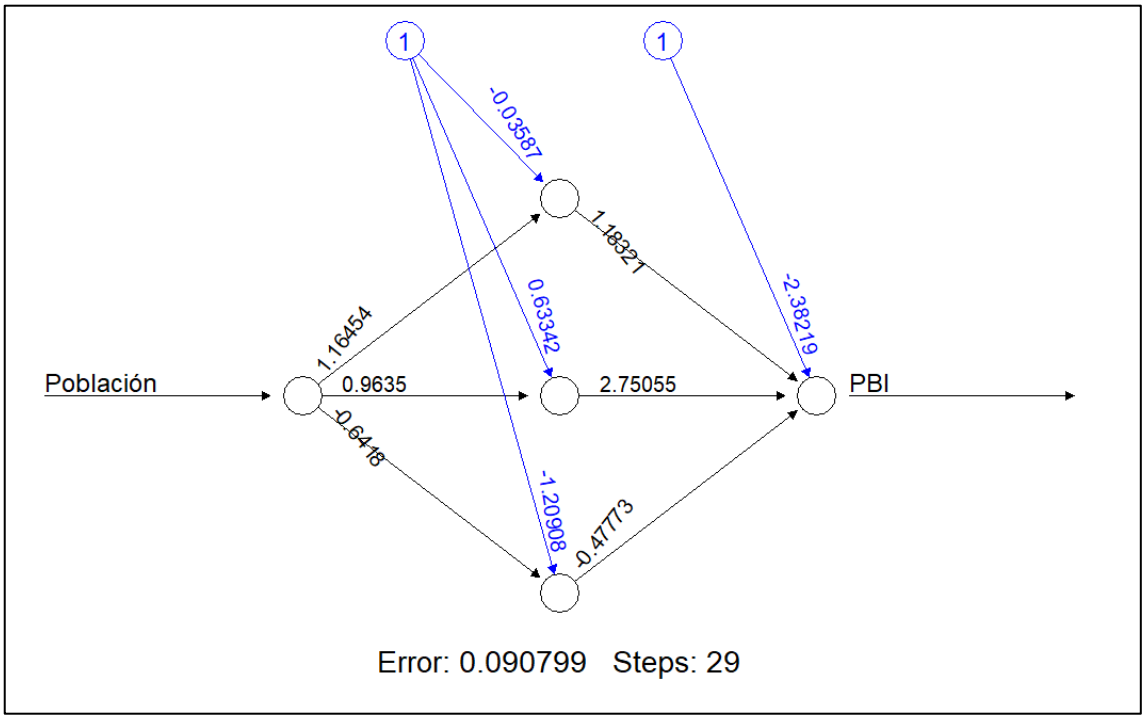
$PBI = 327453 + 825460\ pob + 280969\ pob^2 - 49359\ pob^3 - 134244\ pob^4 - 20204\ pob^5 + 55638\ pob^6 + 19738\ pob^7$

Obtenido el modelo polinómico de orden 7 se verifica el nivel de bondad de su predicción, para lo cual se utilizarán los datos de la Tabla 9 Base de datos de prueba, con los que se calcula el MAPE obteniendo un resultado de 4.776539.

5.4. Red Neuronal

Finalmente se formula el último modelo: la red neuronal de tipo back propagation con una capa oculta conformada por tres nodos, el resultado se representa en la Figura 7

Figura 7 Red neuronal propuesta como modelo rival



Fuente: Elaboración propia

La Red neuronal mostrada en la Figura 7 ha sido entrenada con los datos de la Tabla 8 y, posteriormente, sometida a prueba con los datos de la Tabla 9. Cabe recalcar que este modelo es considerado no paramétrico, por lo que no requiere prueba de supuestos.

Tal como el modelo polinómico, se evaluó la bondad de la red neuronal mediante MAPE cuyo resultado fue 0.1598667.

5.5. Selección del mejor modelo

La selección del mejor modelo se logra evaluando el MAPE y eligiendo a aquel que presente menor MAPE, para lo antes mencionado se presentan los resultados de los modelos en la Tabla 14.

Tabla 14 Comparación de MAPE entre modelos

N. °	Modelo	MAPE
1	Regresión lineal	-
2	Regresión polinómica	4.776539
3	Red neuronal	0.1598667

Fuente: Elaboración propia

Como se puede apreciar en la Tabla 14, el modelo con menor MAPE es la red neuronal.

5.6. Pruebas e hipótesis

- Hipótesis 1: Un modelo basado en la población explica en gran medida la predicción del PBI.

  Como se muestra en la Tabla 12 todos los coeficientes beta relacionados a la población son estadísticamente significativos. Con lo cual existe evidencia a favor de esta hipótesis.

- Hipótesis 2: La regresión polinómica influye favorablemente en la calidad del pronóstico del PBI en función de la población.

  Así también, se observa en el numeral 5.3, que el modelo polinómico de orden 7 es apto para el pronóstico del PBI en función de la población, es necesario tomar en cuenta también que el MAPE de este modelo es significativamente bajo por lo que existe evidencia para esta hipótesis.

- Hipótesis 3: Las redes neuronales influyen favorablemente en la calidad del pronóstico de PBI en función de la población.

  En el numeral 5.4 como en la Tabla 14, se muestra que el MAPE de este modelo es el menor entre los modelos evaluados, por lo tanto, hay evidencia a favor de esta hipótesis.

- Hipótesis general: La regresión polinómica y redes neuronales influyen favorablemente en la calidad del modelo y pronóstico del PBI en función de la población.

  Al haber evidencia a favor de todas las hipótesis secundarias, la hipótesis principal queda demostrada.

5.7. Script utilizado

Se detalla a continuación el script del tratamiento de datos, así como el código para el desarrollo de los modelos lineal, polinómico y de red neuronal.

```
#__________Leyendo datos__________
install.packages("readxl")
library(readxl)
datosPP <- read_excel("C:/Users/micmz/Desktop/Documento
Tesis/DatosPBI94-22.xlsx")
View(datosPP)
```

```
#Graficar los datos
install.packages("ggplot2")
library(ggplot2)
ggplot(datosPP,aes(Población, PBI))+geom_point()+stat_smooth()

#Dividiendo train y test
datos_train=datosPP[1:24,]
datos_test=datosPP[25:29,]
View(datos_train)

#_______Genero modelo lineal_______
modelo_lineal= lm(PBI~Población,datos_train)
summary(modelo_lineal)

#Verificar Normalidad del modelo lineal
shapiro.test(modelo_lineal$residuals)

#Verificar Homocedasticidad del modelo lineal con Breusch-Pagan
install.packages("lmtest")
library(lmtest)
bptest(modelo_lineal)

#Verificar la no existencia de autocorrelación/independencia de errores
#DurbinWatson
install.packages("car")
library(car)
durbinWatsonTest(modelo_lineal)

#Verificar linealidad del modelo/ correcta especificación Prueba RESET
library(lmtest)
resettest(modelo_lineal)
```

```
#_________Genero modelo polinómico__________
#__modelo orden 2_______
modelo_poli2=lm(PBI~poly(Población,2),data = datos_train)
summary(modelo_poli2)
shapiro.test(modelo_poli2$residuals) #prueba de normalidad shapiro wilks
bptest(modelo_poli2) # prueba homocedasticidad Breusch Pagan
durbinWatsonTest(modelo_poli2) #prueba autocorrelacion DurbinWatson
resettest(modelo_poli2) #prueba linealidad

#__modelo orden 3_______
modelo_poli3=lm(PBI~poly(Población,3),data = datos_train)
summary(modelo_poli3)
shapiro.test(modelo_poli3$residuals) #prueba de normalidad shapiro wilks
bptest(modelo_poli3) # prueba homocedasticidad Breusch Pagan
durbinWatsonTest(modelo_poli3) #prueba autocorrelacion DurbinWatson
resettest(modelo_poli3) #prueba linealidad

#__modelo orden 4_______
modelo_poli4=lm(PBI~poly(Población,4),data = datos_train)
summary(modelo_poli4)
shapiro.test(modelo_poli4$residuals) #prueba de normalidad shapiro wilks
bptest(modelo_poli4) # prueba homocedasticidad Breusch Pagan
durbinWatsonTest(modelo_poli4) #prueba autocorrelacion DurbinWatson
resettest(modelo_poli4) #prueba linealidad

#__modelo orden 5_______
modelo_poli5=lm(PBI~poly(Población,5),data = datos_train)
summary(modelo_poli5)
shapiro.test(modelo_poli5$residuals) #prueba de normalidad shapiro wilks
bptest(modelo_poli5) # prueba homocedasticidad Breusch Pagan
durbinWatsonTest(modelo_poli5) #prueba autocorrelacion DurbinWatson
resettest(modelo_poli5) #prueba linealidad
```

```
#__modelo orden 6_______
modelo_poli6=lm(PBI~poly(Población,6),data = datos_train)
summary(modelo_poli6)
shapiro.test(modelo_poli6$residuals) #prueba de normalidad shapiro wilks
bptest(modelo_poli6) # prueba homocedasticidad Breusch Pagan
durbinWatsonTest(modelo_poli6) #prueba autocorrelacion DurbinWatson
resettest(modelo_poli6) #prueba linealidad

#__modelo orden 7_______
modelo_poli7=lm(PBI~poly(Población,7),data = datos_train)
summary(modelo_poli7)
shapiro.test(modelo_poli7$residuals) #prueba de normalidad shapiro wilks
bptest(modelo_poli7) # prueba homocedasticidad Breusch Pagan
durbinWatsonTest(modelo_poli7) #prueba autocorrelacion DurbinWatson
resettest(modelo_poli7) #prueba linealidad

# el modelo polinómico de orden 7 cumple con los supuestos
#__________Se genera una predicción con poli7_________________
predecir = predict(modelo_poli7, datos_prueba)
library(Metrics)
RMSE1=rmse(predecir,datos_prueba$PBI)
install.packages("MLmetrics")
library(MLmetrics)
MAPE1 = MAPE(predecir,datos_prueba$PBI)
bondad = data.frame(RMSE1,MAPE1)
colnames(bondad) = c('RMSE','MAPE')
bondad
#_____________Red Neuronal_____________________


#Estandarizando los datos para el backpropagation
```

```
maxs = apply(datosPP,2,max)
mins = apply(datosPP,2,min)

datos_nrm = as.data.frame(scale(datosPP, center = mins, scale = maxs-mins))
View(datos_nrm) #Revisando los datos

train_nrm = datos_nrm[1:24,]
test_nrm = datos_nrm[25:29,]
#___Genernado red neuronal de 1 capa 3 nodos___
library(neuralnet)
modelo.nn = neuralnet(PBI~Población, data = train_nrm,hidden = c(3), threshold
= 0.05, algorithm = "rprop+" )
plot(modelo.nn)

#Realizando el pronóstico
pr.nn = compute(modelo.nn, within(test_nrm, rm(PBI)))
pr.nn

#volviendo a los datos originales
PBI.predict = pr.nn$net.result*(max(datosPP$PBI) - min(datosPP$PBI)) +
min(datosPP$PBI)

#Calculando el RMSE y MAPE
library(Metrics)
```

```
RMSE2 = rmse(PBI.predict, datos_test$PBI)

library(MLmetrics)

MAPE2 = MAPE(PBI.predict, datos_test$PBI)

bondad2 = data.frame(RMSE2,MAPE2)

colnames(bondad2) = c('RMSE','MAPE')

bondad2
```

CAPÍTULO VI: DISCUSIÓN

Se encuentran similitudes de la presente tesis con la investigación de Castro (2019) en cuanto a la herramienta de medición MAPE para evaluar la calidad de pronóstico obtenido. Así también, como en la presente tesis, la investigación de Espinoza (2022), Chávez (2022) y Quispe (2022) concluyó que el uso de las redes neuronales presenta mejor calidad de pronóstico frente a otros métodos.

Otra similitud es la que se encuentra en la investigación de Espinoza (2022), en donde se mantiene la idea del uso de una arquitectura simple en las redes neuronales. Por otro lado, a diferencia de la presente tesis, en la investigación de Angulo (2013) y Castro (2019) se evaluó al PBI no como resultado de una variable independiente sino como una serie de tiempo.

Otra diferencia se encuentra en la investigación de Angulo (2013) la metodología de Box-Jenkins presenta mejor calidad de pronóstico para la medición del PBI trimestral. Adicionalmente, en la investigación de Castro (2019) no se consideran a las redes neuronales entre las metodologías utilizadas para el pronóstico del PBI.

Por último, a diferencia de la presente tesis Espinoza (2022), Chávez (2022) y Quispe (2022) utilizan una gran cantidad de datos para poder entrenar sus modelos de redes neuronales.

El modelo lineal, primer modelo a evaluar, no obtuvo resultados favorables en las pruebas Durbin Watson de autocorrelación y RESET de modo que fue descartada y no ingresa en la comparativa del MAPE de los tres modelos a evaluar.

La ecuación polinómica de orden 7 fue aquella que obtuvo resultados favorables en las pruebas de Normalidad de residuos, homocedasticidad, Independencia de errores y linealidad del modelo motivo por el cual se eligió para evaluar el modelo polinómico. Este modelo presentó un MAPE de 4.776539.

La red neuronal de una entrada (población), una capa oculta de tres nodos y una salida (PBI) obtuvo un MAPE de 0.1598667, mucho menor al obtenido por el modelo lineal, con lo cual la red neuronal se presenta como el mejor de los tres modelos para pronosticar el indicador macroeconómico PBI con respecto a la población.

La capacidad de pronosticar el PBI puede ser utilizado por las autoridades estatales para la planificación y la asignación de presupuesto a distintos sectores tales como educación o salud. Resultado de lo anterior es útil también para grupos empresariales puesto que la variación del PBI se traduce en una variación de asignación de presupuesto para sectores, lo que a su vez significa más o menos inversión o gasto público, por lo que las empresas podrían realizar una planificación más realista y por ende más eficiente.

Así también, la presente tesis aporta, a la escasa literatura encontrada, un caso más en donde se compara un modelo polinómico y una red neuronal a través de su calidad de pronóstico reflejada en el MAPE.

CAPÍTULO VII: CONCLUSIONES Y RECOMENDACIONES

7.1. Conclusiones

- Existe evidencia para afirmar que la población tiene influencia estadísticamente significativa en el PBI del Perú.
- El modelo lineal no es compatible para el pronóstico del PBI del Perú en función de la población.
- El modelo polinómico de orden 7 influye favorablemente en la calidad de pronóstico del PBI del Perú en función de la población.
- El modelo de red neuronal influye favorablemente en la calidad de pronóstico del PBI del Perú.
- En contraste con el modelo polinómico de orden 7, el modelo de red neuronal ofrece mayor bondad para el pronóstico del PBI del Perú en función de la población.

7.2. Recomendaciones

- Modelar el pronóstico con variaciones de la población tales como la población en edad para trabajar y la población económicamente activa (PEA).
- Modelar el pronóstico de variables distintas a la población que puedan influir en el PBI del Perú.
- Contrastar los resultados obtenidos en esta investigación ante los obtenidos al analizar el pronóstico del PBI como una serie de tiempo.

- Aumentar la cantidad de nodos en la capa de entrada de una red neuronal al añadir variables distintas de la población.
- Ahondar en el estudio sobre una arquitectura definida de red neuronal para una determinada operación.

REFERENCIAS BIBLIOGRÁFICAS

Angulo Elorreaga, L. A. (2013). *Predicción del PBI real nacional trimestral: Redes neuronales autorregresivas vs. Metodología Box-Jenkins.* Trujillo.

Boden, M. (2017). *Inteligencia Artificial.* Madrid: Turner.

Bruce, P., Bruce, A., & Gedeck, P. (2022). *Estadística práctica para ciencia de datos con R y Python.* Marcombo.

Castro Zeña, A. J. (2019). *Modelo de pronóstico del.* Lambayeque.

Chavez Barturén, D. A. (2022). *Aplicación de redes neuronales artificiales para la valorización de los departamentos habitacionales de la ciudad de lima en 2020.* Pimentel.

Cole, M. R. (2018). *Hands-On Neural Network Programming with C#: Add powerful neural network capabilities to your C# enterprise applications.* Birmingham, UK: Packt Publishing Ltd.

Cristancho Giraldo, L. (2023). *Herramientas Para La Medición Económica Y Su Aplicación En Latinoamérica.* ALPHAEDITORIAL. Obtenido de https://www.google.com.pe/books/edition/Herramientas_para_la_medici%C3%B3n_econ%C3%B3mic/fNavEAAAQBAJ?hl=es-419&gbpv=0

El-Sahili, L. F. (2022). *Nuevos modelos integracionistas y cientificistas: Introducción a la psicología.* Iguala de la Indeendencia: Editorial de la Federación Mexicana de Psicología.

Espinoza Quiroz, S. E. (2022). *Diseño de un sensor virtual para la predicción de la eficiencia de conversión de SO_2 a SO_3 en un reactor catalítico mult-ilechos mediante redes neuronales artificiales.* Lima.

Flores López, R., & Fernández Fernández, J. (2008). *Las Redes Neuronales Artificiales.* España: Netbiblo.

Gujarati, D. N. (2018). *Linear regression: a mathematical introduction.* Thousand Oaks: SAGE Publications.

Heredia Alvaro, J. (2001). *Sistema de indicadores para la mejora y el control integrado de la calidad de los procesos.* Castello de la Plana: Publicaciones de la Universitat Jaume I.

Instituto Nacional de Estadística e Informática. (2000). *Metodología del Producto Bruto Interno Anual.*

Instituto Nacional de Estadística e Informática. (2013). *https://www.inei.gob.pe.* Obtenido de https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1178/anexo02.pdf

Instituto Nacional de Estadística e Informática. (agosto de 2022). *https://www.inei.gob.pe.* Obtenido de https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1857/libro.pdf

International Business Machines Corporation. (1 de marzo de 2016). *https://www.ibm.com*. Obtenido de https://www.ibm.com/docs/en/zos-basic-skills?topic=more-what-is-data-set

International Business Machines Corporation. (19 de agosto de 2020). *https://www.ibm.com*. Obtenido de https://www.ibm.com/topics/neural-networks

International Business Machines Corporation. (17 de 08 de 2021). *https://www.ibm.com*. Obtenido de https://www.ibm.com/docs/es/spss-modeler/saas?topic=nodes-statistical-models

Isasi Viñuela, P., & Galván León, I. M. (2004). *Redes neuronales artificiales Un enfoque práctico.* Madrid: Pearson Educación S.A.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Nueva York: Farrar, Straus and Giroux.

Malthus, T. (1846). *An Essay on the Principle of Population.* (J. Noguera, & J. Miquel, Trads.) Madrid: Lucas Gonzalez.

Mankiw, G. N. (2017). *Principios de la economía.* México: Cengage.

Marquina Vila, A. (2017). *Elogio del algoritmo: Las matemáticas del cálculo científico.* Valencia: Universitat de Valencia.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis.* New Jersey: John Wiley & Sons, Inc.

Müller-Esterl, W. (2008). *Bioquímica: fundamentos para medicina y ciencias de la vida.* Barcelona: Reverté.

Pajares Mantinsanz, G., & De la Cruz García, J. (2022). *Aprendizaje Automático.* Madrid: Ediciones de la U.

Peng, M. (2018). *Negocios globales.* Ciudad de México: Cengage.

Pérez Borrero, I., & Gegúndez Arias, M. E. (2021). *DEEP LEARNING.* Huelva, España.

Perez, I., & Gegúndez, M. E. (2021). *DEEP LEARNING.* Huelva: Publicaciones de la universidad de Huelva.

Perez, R., Caso, C., Río, M., & López, A. J. (2011). *Introducción a la Estadística Económica.* Oviedo, España: Universidad de Oviedo.

Priyadarshini Jeyaraman, B., Renbo Olsen, L., & Wambugu, M. (2019). *Practical Machine Learning with R: Define, build, and evaluate machine learning models for real-world applications.* Birmingham: Packt Publishing Ltd.

Quispe Mamani, N. (2022). *Implementación de redes neuronales artificiales para regresión en predicción de volumen de pedidos de textiles de moda.* Lima.

R4V Plataforma de Coordinación interagencial para Refugiados y Migrantes de Venezuela. (12 de octubre de 2022). *https://www.r4v.info.* Obtenido de https://www.r4v.info/es/document/r4v-america-latina-y-el-caribe-refugiados-y-migrantes-venezolanos-en-la-region-sept-2022

Valbuena, R. (2021). *Inteligencia Artificial: Investigación Científica Avanzada Centrada en Datos.* Maracaibo: Centro de Capacitación Alternativa CENCAL.

Vásquez Conde, R. (2020). *Biología.* Patria Educación.

Villamizar Jaimes, M. (2021). *Macroeconomía.* Colombia: Ediciones de la U.

Vinuesa, P. (22 de Octubre de 2016). *https://www.ccg.unam.mx.* Obtenido de https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema9_regresion.html#tipos-de-regresion