

## **Algoritmo genético y red neuronal artificial para la optimización de portafolios de inversión**

### **Genetic algorithm and artificial neural network for optimization of investment portfolios**

Bach. Gustavo Alexis Flores Fernandez

Ingeniero Industrial, Universidad Nacional de Piura ORCID [0000-0002-4488-4325](https://orcid.org/0000-0002-4488-4325)

Mgtr. Erick Chuquitapa Rojas

Magister en Políticas Públicas, Universidad Nacional de San Antonio Abad del Cusco

[0000-0003-2635-1389](https://orcid.org/0000-0003-2635-1389)

Dr. Miguel Jiménez Carrión

Ingeniero Industrial, Universidad Nacional De Piura, ORCID [0000-0001-9632-5085](https://orcid.org/0000-0001-9632-5085)

### **Resumen**

La presente investigación tuvo como objetivo construir un algoritmo genético y red neuronal artificial para optimizar portafolios de inversión, considerando que en la teoría moderna de portafolios de inversión, la optimización es un problema multiobjetivo que contempla la maximización del rendimiento y la minimización de la volatilidad o también llamado riesgo, esto abre la posibilidad de un espacio de soluciones altamente combinatorio, convirtiéndose en un problema con complejidad computacional que no puede ser resuelto por algoritmos determinísticos. Para poder cumplir con el objetivo, se evaluaron 255 empresas que realizan sus actividades dentro del ámbito nacional peruano y cotizan sus acciones en la Bolsa de Valores de Lima. En esta investigación obtuvieron como resultados un error cuadrático medio de 6.33%, un error absoluto medio de 5.07% y una precisión de 92.35% relacionada con la red neuronal artificial lo cual evidencia una aceptable capacidad de generalización para predecir la tendencia positiva de las acciones que serán entradas para el algoritmo genético. Respecto al algoritmo genético se logró modelar la función de calidad que consideró 5 factores relacionados a la rentabilidad y la volatilidad de las acciones, así como la diversificación del portafolio, finalmente se encontró como mejor configuración del algoritmo genético la arquitectura que tuvo como fitness un valor de 0.772482, que

se traduce en una rentabilidad de 1.00058% lo que se traduce en 12.00696% anual; y volatilidad de 0.00612%. Se concluye que el algoritmo genético optimiza portafolios de inversión al encontrarse una rentabilidad mayor y una volatilidad menor comparada con otros métodos siendo específicamente la volatilidad un porcentaje mucho menor.

**Keywords:** metaheurísticas, algoritmos genéticos, red neuronal artificial, optimización, portafolios de inversión

## 1 Introducción

Las unidades productivas que dinamizan las economías locales y nacionales, tienen como principal necesidad, la planificación en sus inversiones, para adecuarse mejor a su demanda; esto implica generalmente a evaluar en un horizonte de planeamiento, el futuro escalamiento de infraestructura u operaciones, mediante una evaluación de proyectos de inversión (Baca, 2016). Muchas veces la evaluación de proyectos de inversión no contempla dentro de su estado de resultados proyectados, escenarios realistas y detalles específicos propios de las operaciones de las unidades productoras, como son la adquisición de activos fijos en los periodos transitorios de escalabilidad empresarial, inversiones no tangibles, etc. Tal como menciona Hart & Zingales (2011), la necesidad de invertir en activos, que muchas veces no son contemplados en los estados financieros, evidencian la falta de herramientas para mejorar la eficiencia de las inversiones (p.4.), evidenciando el desaprovechamiento del exceso de liquidez en la unidad productora, que conllevaría a que los flujos netos acumulados reflejen de forma más adecuada el dinamismo empresarial, y no caer en rendimientos negativos sin haber intentado mejorar la eficiencia en inversiones externas.

El proyectista que se enfrente a un exceso de liquidez o riqueza en sus flujos proyectados, debe tomar una decisión acertada para mejorar la eficiencia del proyecto, y ello conlleva, decidir, si toma los costos de oportunidad, entre los cuales se puede contemplar el mantener el dinero, evitando o asumiendo los cambios inflacionarios, o hacer una inversión libre de riesgo invirtiendo en una entidad financiera, o en caso contrario, según lo que indica Ames (2012), invirtiendo en el mercado de capitales, lo cual es una opción viable para los jefes de proyectos que asumen un riesgo,

y en donde encontrarán un refugio de valor con esto se puede tener como consecuencia un rendimiento mayor a la inversión en una entidad financiera.

Frente a los retos que enfrenta el proyectista se debe tomar la mejor opción para invertir y esa decisión es un problema NP-completo, tal como lo manifiesta Didier et al. (2021) cuando afirma que la opción de generar un rendimiento en el mercado de capitales, considerando que existe un riesgo, por lo tanto, es de vital importancia escoger de manera eficiente los activos o evaluar la cartera de inversión óptima, y de forma consecuente se puede proyectar un rendimiento mayor en la inversión, y minimizar el riesgo asumible relacionado con la volatilidad de los mismos activos (p.2). Considerando que cada activo de capital tendrá un porcentaje de inversión de los rendimientos de exceso de los flujos proyectados, se tendría un proceso altamente combinatorio de opciones, por ejemplo si se evaluaran 5 activos, se tendría 10,000,000,000 combinaciones, siendo las representaciones porcentuales, además si cada combinación se evaluará en una función objetivo con una demora de tiempo de 1 segundo, se tendría un tiempo exponencial en poder evaluar todas las combinaciones para 10 o 20 activos. Por lo tanto, este problema con dos objetivos contrapuestos; por ser de maximización en el caso del rendimiento, y de minimización en el caso del riesgo, no puede ser resuelto en todas sus combinaciones en un tiempo razonable mediante un algoritmo determinista, y considerando su naturaleza combinatoria se categoriza finalmente como un problema “No polinomial” (Gutiérrez et al., 2020). Si bien es cierto que en la evaluación de portafolios de inversión se tiende a realizar la evaluación de la cartera con una cierta cantidad de combinaciones para representar la frontera de Pareto y utilizar el índice de Sharpe para encontrar el óptimo, no existe una solución determinista que contemple y evalúe de forma iterativa una gran cantidad de soluciones eficientes y mucho menos que contemple la optimización de los dos objetivos al mismo tiempo para encontrar soluciones factibles. Por lo que se hace de vital importancia el uso de herramientas de inteligencia artificial como son las redes neuronales artificiales y metaheurísticas como los algoritmos genéticos, tal como indican Das et al. (2023) y Li et al. (2023), quienes han demostrado una utilidad fundamental en el campo de las finanzas, debido a su efectividad de predicción; y a la capacidad de hacer iteraciones exhaustivas en los espacios de búsqueda, mediante la exploración y explotación de individuos factibles, encontrándose soluciones óptimas a problemas complejos mediante la definición correcta de hiperparámetros, como lo son las probabilidades de cruzamiento, mutación y número de

generaciones que sirven para controlar el comportamiento y eficiencia del algoritmo, permitiendo encontrar individuos óptimos en un tiempo razonable.

Como antecedentes, autores como Qi Li et al. (2023) centran su investigación en presentar y desarrollar un enfoque mejorado que combina el Algoritmo Genético Simbólico (SGA) con la Red Neuronal de Memoria a Corto y Largo Plazo (LSTM) para anticipar los rendimientos de las acciones en el mercado chino. Como resultados se obtuvo un rango entre 15.26% a 22,35% de rentabilidad anual.

Por otro lado, Bo (2023) tuvo como objetivo optimizar la composición de la cartera, reducir riesgos e incrementar los rendimientos potenciales mediante el análisis de interdependencias y correlaciones entre activos financieros utilizando redes complejas, y también empleando algoritmos genéticos como técnica de optimización. Como resultados el autor obtuvo un rendimiento del 0.2667 y un radio de Sharpe de 0.0685

Vasiani et al. (2020) buscaron optimizar las carteras de acciones utilizando el método del índice de prioridad y algoritmos genéticos. Su metodología implica seleccionar acciones basadas en el índice de prioridad, considerando parámetros como la ratio precio/beneficio (P/E), beneficio por acción (EPS), creación de riqueza, infravaloración y ratio precio/beneficio-crecimiento (PEG). Las acciones en cada sector se eligen en función de un puntaje de índice de prioridad igual o mayor que el puntaje mínimo de las acciones seleccionadas. Como resultados se obtuvo un máximo de rentabilidad de 14.08% sin embargo se debe tener en cuenta que solo se considera como métrica principal el rendimiento.

Lim et al. (2020) formularon el objetivo principal de diseñar una cartera óptima utilizando un Algoritmo Genético (GA) que incorpora estrategias de impulso y valoración de activos. Su metodología implica analizar los rendimientos ajustados al riesgo en períodos anteriores, utilizando el impulso de estos rendimientos como impulso para la selección de acciones. Sin embargo, reconocen que los movimientos históricos por sí solos no son suficientes para predecir cambios futuros o garantizar rendimientos positivos. Por otro lado, los autores no consideraron el uso de técnicas de inteligencia artificial especializada como son las redes LSTM

Candia et al. (2020) abordaron el problema de la selección de cartera de proyectos para la adjudicación de obras públicas mediante concursos de méritos abiertos supervisados por el

Instituto Nacional de Vías (INVIAS) en Colombia. Metodológicamente, evaluaron dos enfoques alternativos: un algoritmo genético meta-optimizado (GA) cuyo fitness promedio fue de 0.21748 con un tiempo de ejecución de 8.102 minutos superando al procedimiento de búsqueda codicioso probabilístico adaptativo meta-optimizado (GRASP).

Rodríguez et al. (2020) propusieron un método alternativo utilizando algoritmos evolutivos, específicamente un Algoritmo Genético Canónico, para diseñar una cartera de inversiones en divisas llamada "cartera de divisas". Descriptivamente, seleccionaron seis monedas en relación con el peso mexicano, guaraní paraguayo, peso uruguayo, boliviano, dólar estadounidense, libra esterlina y euro. Los autores obtuvieron como resultado un rendimiento de 0.049202% y una volatilidad de 0.003543%

Maholi et al. (2019) tuvieron como objetivo predecir los valores futuros de las acciones utilizando redes neuronales artificiales (ANN) y luego utilizar un algoritmo genético (GA) para formar carteras óptimas que maximicen los rendimientos y minimicen el riesgo. Su metodología comienza aplicando ANN, un modelo de aprendizaje automático representado por una red neuronal densa, lo cual cabe recalcar no son redes especializadas para la evaluación de series de tiempo. Como resultados se obtuvo para un error cuadrático medio de 5.60% pero con un solo paso en el tiempo, además en cuanto a la rentabilidad se obtuvo un 1,42% pero considerando una volatilidad del 0.15%.

Liagkouras, K. (2019) tuvo como objetivo abordar las limitaciones de las técnicas de algoritmos evolutivos existentes en la resolución de problemas combinatorios a gran escala debido a su extenso espacio de búsqueda. Su metodología implica probar el rendimiento del algoritmo propuesto en la asignación óptima de recursos limitados a una serie de oportunidades de inversión competitivas para optimizar objetivos. Como resultados en el análisis experimental obtuvieron un rendimiento no mayor al 0.785 y una volatilidad del 0.0097.

## **2 Metodología**

Los pasos para el desarrollo de la presente investigación consistieron en: 1) Obtener información de todas las empresas listadas en la Bolsa de Valores de Lima las cuales fueron 255, cuya data es de acceso abierto y fue obtenida mediante la página web de la institución

mediante archivos CSV, 2) Realizar un análisis exploratorio de datos, el cual baso en una preprocesamiento mediante la identificación de outliers e imputación de datos mediante una interpolación lineal, 3) Diseñar una red neuronal artificial de Memoria a Corto y Largo Plazo (LSTM) para predecir la tendencia de los valores bursátiles de las empresas, 4) Diseñar un algoritmo genético para optimizar el portafolio de inversiones, a partir de las predicciones de la red neuronal LSTM 5) Desarrollar la discusión de los resultados y 6) Definir las conclusiones basadas en los resultados.(Figura 1)

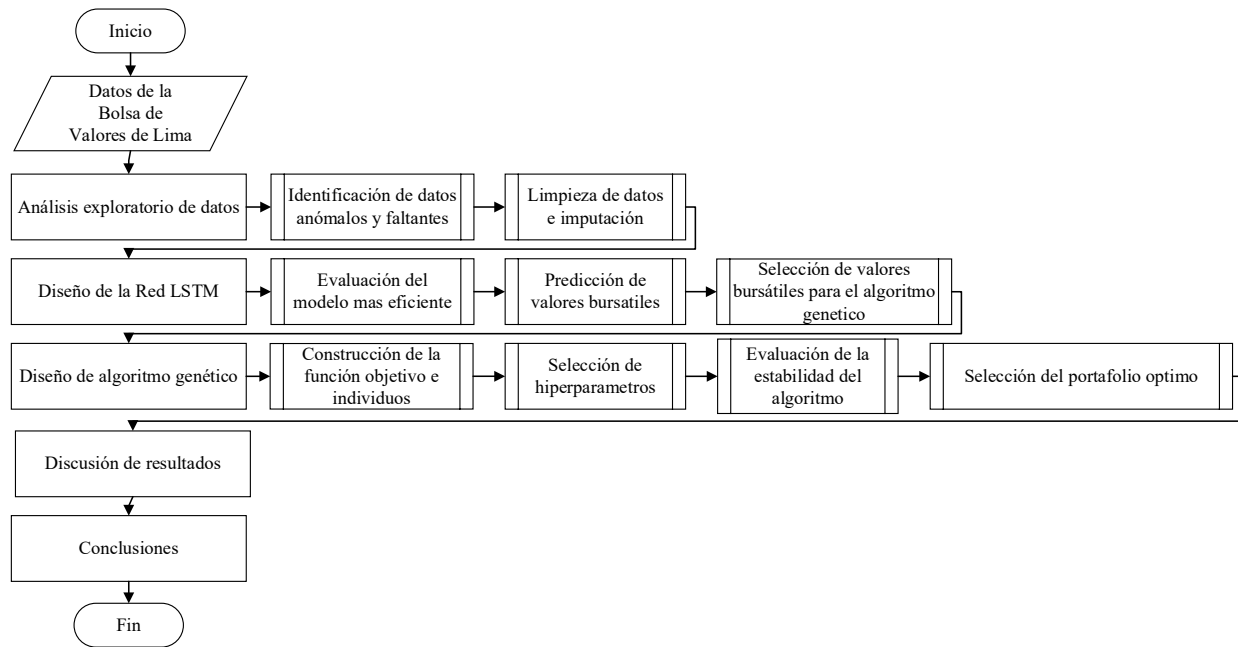


Figure 1. Diagrama de flujo de la metodología aplicada

### 3. RESULTADOS

#### Análisis exploratorio de datos de la Bolsa de Valores de Lima

Para obtener entradas para el modelo de red neuronal, se desarrolló un preprocesamiento de datos para estandarizar los datos y hacerlos útiles para que la red genere datos consistentes en su proceso de aprendizaje y uso general. Por lo tanto, el proceso se guía por el procedimiento descrito por Ruochen & Muchao (2021), comenzando con el manejo de datos faltantes, que es una característica fundamental para que sean entradas de la red neuronal. Para el proceso de adquisición de datos, se recolectaron datos en un solo punto de tiempo, y los activos fueron descartados gradualmente en función de los datos que presentaban y se filtraron. Inicialmente, el portal de la BVL enumeró registros y datos para 255 empresas, de las cuales solo el 85%

(216) proporcionó información sobre acciones. Al explorar los archivos CSV, se descubrió que solo el 57% del total (145) de las empresas tenían datos históricos, mientras que el resto solo proporcionaba la fecha en que se debían haber registrado los parámetros del comportamiento del mercado. A continuación, se determinó el rango de tiempo más amplio de los datos, es decir, los límites inferiores y superiores. Se encontró que las empresas mostraban datos a partir del 2 de enero de 2012, lo que representaba el 38.43% (98) de las empresas, y el 18% (46) tenía datos de cotización de años anteriores, por lo que la fecha anterior se estableció como límite inferior. Con respecto al límite superior, como el año de investigación aún no había terminado, se consideró el 2 de enero de 2023 para asegurar la simetría en los períodos completos. Posteriormente, se evaluó cada registro, revelando que 85 de las empresas filtradas tenían un rango máximo de datos de 2764 registros, de las cuales solo 13 de las empresas tenían fechas faltantes. Luego se evaluaron las empresas con esta cantidad de registros y se encontró que faltaban datos de parámetros de activos. Luego se evaluó el porcentaje de datos faltantes para establecer una tolerancia que no afectara el comportamiento del modelo en diseño.

Según la tabla 1, en función de los 9 parámetros que exponen el comportamiento de las acciones, que fueron, precio de apertura, precio de cierre, precio máximo, precio mínimo, precio promedio, cantidad negociada, monto negociado, fecha anterior, cierre anterior corregido (precio de cierre de una acción ajustado para tener en cuenta eventos que afectan el precio de las acciones), se agrupó en 10 rangos que representan la cantidad de datos faltantes, esto con el fin de conocer el porcentaje de datos faltantes en el total de las 86 empresas evaluadas, y mantener la consistencia con respecto a la cantidad de datos faltantes de conformidad con los autores como Dagnino (2014) y Bennett (2001), quienes recomiendan un margen de tolerancia admisible del 10%, considerando que la ausencia de un porcentaje mayor puede generar sesgos o la pérdida de la distribución natural de los datos; esto se representa en la tabla 1 donde se presentan 10 rangos, donde cada uno representa un intervalo del 10% de datos faltantes, es decir, en R1 están las empresas que tienen entre 0 y 10% de datos faltantes, en este caso por ejemplo el parámetro "Cantidad negociada" el 12% significa que hay 12% de empresas que tienen entre 0 y 10% de datos faltantes en ese parámetro y 58% de empresas que tienen datos faltantes entre 0 a 10% en el parámetro "Fecha anterior" y así sucesivamente.

Table 1. Representación porcentual de empresas según rango de datos faltantes

Porcentaje de empresas según la cantidad de datos faltantes									
Rango de datos	Apertura	Cierre	Máximo	Mínimo	Promedio	Cantidad negociada	Monto negociado (S/)	Fecha anterior	Cierre anterior corregido
R <sub>1</sub>	8%	8%	8%	8%	12%	12%	12%	58%	59%
R <sub>2</sub>	6%	6%	6%	6%	4%	4%	4%	19%	19%
R <sub>3</sub>	4%	4%	4%	4%	6%	6%	6%	9%	11%
R <sub>4</sub>	4%	4%	4%	4%	1%	1%	1%	5%	4%
R <sub>5</sub>	0%	0%	0%	0%	5%	5%	5%	6%	6%
R <sub>6</sub>	6%	6%	6%	6%	5%	5%	5%	0%	0%
R <sub>7</sub>	8%	8%	8%	8%	6%	6%	6%	0%	0%
R <sub>8</sub>	8%	8%	8%	8%	11%	11%	11%	1%	1%
R <sub>9</sub>	13%	13%	13%	13%	11%	11%	11%	1%	0%
R <sub>10</sub>	45%	45%	45%	45%	42%	42%	42%	2%	2%

Considerando que el porcentaje máximo admisible de datos faltantes es 10%, solo se tomara en cuenta las empresas del rango 1, y es en este rango donde solo el parámetro de cierre anterior corregido tiene un porcentaje representativo del 59%, lo que representa 50 empresas en el primer intervalo, cuyos datos faltantes son inferiores al 10% específicamente 24 presentan datos faltantes y 26 presentas todo el registro de datos. Por lo tanto, se determina que el parámetro a utilizar para generar la proyección de la red neuronal será el cierre anterior corregido, definiendo el uso de un modelo de red neuronal recurrente LSTM multivariable.

Dentro de los datos, no se encontraron valores atípicos; sin embargo, se identificaron 24 empresas con datos faltantes. Para esto, se evaluó la imputación utilizando la interpolación lineal y también el método K-means, que es un algoritmo de aprendizaje automático que, para el caso específico de la imputación, genera un punto de datos basado en los k vecinos más cercanos al punto de datos faltante, como se recomienda en la revisión de Fang & Wang (2020).

Para evaluar cual fue el mejor método de imputación se hizo una comparación entre los dos métodos usados, para conocer cual tiene la mayor cantidad de empresas con mejor coeficiente R<sup>2</sup> (Coeficiente de determinación)

Los resultados después de la imputación en relación con R<sup>2</sup> se muestran en la tabla 2.



Tabla 2. Comparación de los coeficientes de determinación de los métodos de imputación de datos

<b>Nemónico</b>	<b>Modelo</b>	
	<b>KNN</b>	<b>Interpolación</b>
CARTAVC1	0.0293	0.0219
BUENAVC1	0.2269	0.2241
BACKUAC1	0.4843	0.5402
AIHC1	0.6604	0.6868
FALABEC1	0.5235	0.5271
EXALMC1	0.0598	0.0601
ELCOMEI1	0.8295	0.8646
CREDITC1	0.462	0.4627
LAREDOC1	0.8067	0.858
HIDRA2C1	0.7759	0.7802
GLORIAI1	0.3341	0.341
PERUBAI1	0.1559	0.1594
PHTBC1	0.4889	0.474
POSITIC1	0.1931	0.1992
SNJACIC1	0.1185	0.1139
TUMANC1	0.4978	0.4959
TEF	0.8633	0.869
SPCCPI1	0.0015	0.0001
SIDERC1	0.4874	0.4878
EGEPIBC1	0.0015	0.0002
PODERC1	0.6065	0.7313
SAGAC1	0.4295	0.4173
HIDROSI1	0.6927	0.7642
GBVLAC1	0.6884	0.8218

Porcentaje de representación de empresas con un coeficiente de determinación mayor	33%	67%
--	-----	-----

Teniendo en cuenta que, en el modelo de interpolación, el 67% de las empresas muestra un coeficiente de determinación más alto, este modelo fue elegido como el método de imputación.

**Diseño de una red neuronal para la predicción de acciones como entrada para el algoritmo genético**

Para predecir el precio de cierre de la acción, que, debido a sus características de utilizar un estado previo y poder generar el siguiente estado, es adecuada para modelos de series temporales donde han demostrado su utilidad en múltiples tareas de inteligencia artificial (Hewamalage et al., 2021). En el caso de la Bolsa de Valores de Lima, la herramienta utilizada es el promedio móvil a través del análisis técnico, según lo confirmado al acceder al portal oficial de esta entidad. En el caso del presente estudio, el diseño se propuso basado en el tipo de redes neuronales recurrentes llamadas Memoria a Corto y Largo Plazo (LSTM), que tienen la capacidad de recordar a través de puertas de pérdida de memoria, que, mediante funciones no lineales, eliminan o preservan información, puertas de entrada que deciden qué información es aceptada por el modelo, y puertas de salida donde se decide qué parte de la memoria LSTM contribuye a la salida (Siarni-Namini et al., 2019).

La estructura de una red LSTM para este caso considera el valor de cierre corregido de la serie temporal de valoración de acciones, como se mencionó en el análisis exploratorio, y su salida será una medida de tendencia definida bajo el concepto de que la tendencia es alcista si su valor de cotización es mayor que en el período anterior. Sin embargo, esta expresión debe complementarse con una medida de holgura para mejorar la precisión del límite que puede aceptarse como valor de tendencia alcista, basado en los datos analizados. Por lo tanto, se determinó que la expresión se complementaría según la ecuación 1:

$$P_{(t+1)} > P_{(t)} + [\sigma_{\%Mercado}] * P_{(t)} \tag{Ec. 1}$$

Donde:

- $P_{(t)}$ : Precio de cierre de la acción en el periodo  $t$
- $P_{(t+21)}$ : Precio de cierre de la acción en el periodo  $t + 21$
- $\sigma_{\%Mercado}$ : Representación porcentual de la desviación estandar del mercado

Según la expresión anterior, se utiliza la medida de desviación estándar del mercado para que la tendencia tenga un excedente que absorba cualquier variación que exista debido a alguna variable atípica dentro del mercado internacional, considerando que entre el 20% y el 50% del comportamiento de los precios se genera a partir de fuerzas del mercado (Gitman & Joehnk, 2009, p. 357). Por lo tanto, insertar en la ecuación una medida que considere las fluctuaciones que el comportamiento general de las acciones ha estado experimentando ayuda a mitigar cualquier condición que no garantice un rendimiento superior en un marco de tiempo posterior, que también se define después del día 21 como la rentabilidad esperada a corto plazo (Cáceres, 2018). De manera similar, en cuanto al indicador utilizado para determinar esta medida de desviación, se utilizó el Standard & Poor's 500 (S&P500), que se define como un índice auténtico al considerar la capitalización de las 500 empresas más grandes de Estados Unidos y capturar el 80% de toda la capitalización.

### Algoritmo Univariable-Multistep

Basándose en el enfoque previamente formulado y mostrado en el Algoritmo 1, se propusieron cuatro topologías de redes LSTM para su evaluación. Los hiperparámetros considerados incluyen la entrada de la red, que consiste en lotes de 42 días, y una salida de 21 días, que representa el valor de cierre de la acción para cada empresa tal como se muestra en la tabla 3. Además, se consideraron el número de neuronas en cada capa de la red para cada topología, el número de épocas y la tasa de aprendizaje, como se muestra en la siguiente Tabla 4.

Tabla 3. Valor de cierres de las acciones para cada empresa

	Nemónico							
Fecha	CASAGRC1	CARTAVC1	.	.	.	SAGAC1	HIDROSI1	GBVLAC1

---

							S/	S/	S/
2/01/2012	S/	15.15	S/	23.00	.	.	4.25	0.60	13.50
							S/	S/	S/
3/01/2012	S/	15.20	S/	23.00	.	.	4.25	0.60	13.50
							S/	S/	S/
4/01/2012	S/	15.30	S/	23.00	.	.	4.25	0.60	13.50
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
							S/	S/	S/
29/12/2022	S/	8.00	S/	35.00	.	.	8.97	0.22	2.45
							S/	S/	S/
30/12/2022	S/	7.85	S/	35.10	.	.	8.97	0.22	2.60
							S/	S/	S/
2/01/2023	S/	7.90	S/	35.10	.	.	8.97	0.22	2.60

---

---

**Algorithm 1** Modelo de Red LSTM Univariado-Multistep

---

```
1: Entrada: Librerías
2: Entrada: Acciones del análisis exploratorio
3: Entrada: hiperparametros
4: Salida: Métricas de rendimiento del algoritmo
5: Data frame  $\leftarrow$  Valores de las acciones
6: tr, vl, tst  $\leftarrow$  set de entrenamiento, validación y test
7: function ESCALAMIENTO DE DATOS
8:   Set de entrenamiento escalado  $\leftarrow$  Escalador MinMax Scaler(Set de entrenamiento)
9:   Set de validación escalado  $\leftarrow$  MinMax Scaler(Set de validación)
10:  Set de prueba escalado  $\leftarrow$  MinMax Scaler(Set de prueba)
11: fin function
12: Para empresas del data frame hacer
13:   function ENTRENAR MODELO
14:     Modelo  $\leftarrow$  Secuencial
15:     Modelo  $\leftarrow$  agregar capa LSTM
16:     Modelo  $\leftarrow$  agregar capa Densa
17:   fin function
18:   function METRICAS DE RENDIMIENTO
19:     RMSE, MAE Precisión  $\leftarrow$  Desempeño del modelo
20:   fin function
21: fin Para
22: Salida: Métricas del rendimiento del modelo
```

---

Basándonos en los tres parámetros considerados para la evaluación de la red, tenemos el error cuadrático medio, el error absoluto medio y la precisión. Después de entrenar la red, se obtuvieron los siguientes resultados, como se muestra en la tabla 5.

Tabla 4. Comparación de modelos de red LSTM

<b>Características del modelo</b>	<b>Modelo 1</b>	<b>Modelo 2</b>	<b>Modelo 3</b>	<b>Modelo 4</b>
Número de unidades de procesamiento	50	100	150	200
Épocas	200	300	400	500
Factor de aprendizaje	0.0002	0.0003	0.0004	0.0005
Tiempo de ejecución optimizado con GPU	35 minutos	31 minutos	33 minutos	58 minutos

Topología de la red

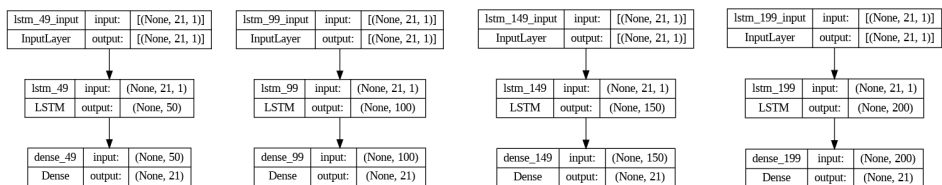


Tabla 5. Comparación de las métricas de los modelos

Métrica	Modelo 1	Modelo 2	Modelo 3	Modelo 4
RMSE_train	3.85%	3.75%	3.69%	3.57%
RMSE_val	5.81%	5.47%	5.37%	5.16%
RMSE_test	7.25%	6.77%	6.76%	6.33%
MAE_train	3.09%	2.97%	2.89%	2.75%
MAE_val	4.71%	4.33%	4.19%	3.94%
MAE_test	6.21%	5.67%	5.56%	5.07%
Accuracy_train	96.58%	96.85%	96.90%	97.08%
Accuracy_val	96.81%	96.56%	97.00%	97.14%
Accuracy_test	90.46%	91.21%	91.98%	92.35%

Como se observa en la tabla anterior, el Modelo 4 logra un error cuadrático medio más bajo en comparación con los otros modelos en su conjunto de pruebas, donde se probó la capacidad de generalización de la red con datos teniendo un comportamiento eficiente con datos que nunca ha visto. Además, se consideró el error absoluto medio y se comparó la métrica de precisión con los otros modelos, resultando en una precisión del 92.35%. Por lo tanto, se elige el Modelo 4 para evaluar la tendencia de los precios de las acciones de las empresas que servirán como entrada para el algoritmo genético.

Habiendo identificado la arquitectura de red neuronal LSTM más eficiente, procedimos a utilizar la ecuación 1, la cual al ser aplicada se determina que solo 24 activos de capital cumplen con la condición que considera a empresas con una tendencia al alza en un periodo de 21 días y que absorben las variaciones el mercado, las cuales mejorarán el rendimiento del portafolio de inversión. Por lo tanto, estos activos servirán como entradas para el algoritmo genético.

### **Modelado de los elementos del algoritmo genético**

La estructura del algoritmo genético debe incluir la estandarización de datos para permitir el procesamiento, comenzando desde la estructura de individuos o soluciones potenciales a través

de la representación cromosómica, así como su función de aptitud que evaluará la calidad de cada individuo para que cada uno de los candidatos con mejores características genere descendencia y se generen mejores soluciones.

Para esta sección, se considera que hay 24 activos de capital, por lo que el espacio de muestra se denota de la siguiente manera:

$$\Omega = \{S \subseteq \{A_1, A_2, \dots, A_{24}\} / |S| \geq 1\} \quad \text{Ec. 2}$$

Donde:

$\Omega$ : Espacio muestral de todos los activos de capital

S: es un subconjunto de los activos disponibles

$|S|$ : Restricción que asegura que cada portafolio tenga al menos un activo

Para la optimización del algoritmo genético, cada individuo representa una solución al problema propuesto, lo que significa que cada cromosoma contiene los porcentajes de capital que se distribuirán entre cada uno de los activos. Para esta investigación, se han identificado 24 activos de capital para su evaluación dentro del algoritmo genético; por lo tanto, los individuos se representarán mediante un vector con porcentajes.

Su representación es la siguiente:

$$C = [w_1, w_2, \dots, w_n] \quad \text{Ec. 3}$$

Donde:

$n$ : es el número de activos

$w_i$  : el peso o representación porcentual de inversión asignado al activo  $i$

Este vector  $C$  debe estar sujeto a las restricciones:

$$0 \leq \omega_i \leq 1 \text{ para todo } i$$

$$\sum_{i=0}^{i=n} \omega_i = 1 \quad \text{Ec. 4}$$

Por lo tanto, los individuos tendrán la siguiente forma cromosómica en donde cada peso asignado será un gen, esto se observa:

$W_1$	$W_2$	$W_3$	$W_4$	$W_5$	.	.	.	$W_{24}$
-------	-------	-------	-------	-------	---	---	---	----------

Por lo tanto, el genotipo serán números porcentuales y el fenotipo tendrá la misma representación.

### Modelado de la función objetivo



La función objetivo debe incorporar múltiples elementos o factores que representan el objetivo de optimización, ya que las teorías existentes sugieren que existen factores sistemáticos u otros que ayudan a generar un comportamiento de rendimiento y volatilidad con mayor precisión. Para este estudio en particular, los factores considerados son el rendimiento, la volatilidad, el coeficiente de Sharpe, la prima de riesgo de mercado y la diversificación del portafolio, considerando que los inversionistas tendrán una predisposición o confianza en un portafolio que genere mayores rendimientos por cada unidad de riesgo.

La función objetivo se construye teniendo en cuenta los siguientes elementos:

$$F. O=Fitness = Max (Rentabilidad) + Min (Volatilidad)= MAX \sum \lambda_i F_j \quad \text{Ec. 5}$$

Donde:

$F_j$ : Es el factor  $j$  considerado dentro de la función objetivo

$\lambda_i$  : es el peso  $i$  asignado a cada factor de la función objetivo

Como se ha mencionado anteriormente el acceso a la información se limita a los datos del comportamiento de los activos, por lo que los factores rentabilidad y volatilidad estarán presentes dentro de la ecuación como factores relevantes expresadas de la siguiente manera:

$\lambda_1 R_p$ : Rendimiento del portafolio

$-\lambda_2 \sigma$ : Volatilidad del portafolio que en este caso es negativa debido a que castiga a la función objetivo considerando que el inversor busca minimizarla

Así también otros factores que se consideraron, es la orientación del inversor por asignar recursos a portafolios que maximicen el rendimiento por cada unidad de riesgo invertido, siendo este indicador el radio de Sharpe:

$$\lambda_3 \left( \frac{R_p - R_f}{\sigma} \right) \quad \text{Ec. 6}$$

Se debe considerar para el caso de este factor una restricción específica considerando que solo se debe tener en cuenta este factor si es el que el rendimiento del portafolio es mayor al de la tasa libre de riesgo:

Si  $R_p \geq R_f$  entonces:  $\lambda_3 \left( \frac{R_p - R_f}{\sigma} \right)$

Sino:  $\lambda_3 = 0$

Otro factor que se consideró relevante es la prima de riesgo sobre el mercado, en donde para este elemento al no tener una formula conocida se consideró usar una función sigmoide teniendo en cuenta que mientras mayor sea la diferencia entre el rendimiento del portafolio y del mercado, el valor de esta función se orientara a 1 y mientras menor se aproximara a su valor asintótico 0 tal como se muestra en la siguiente figura 2:

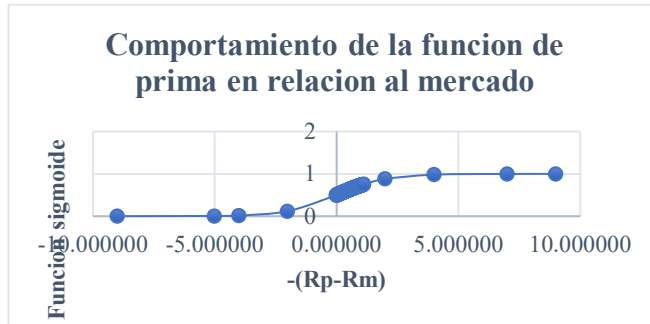


Figura 2. Representación de la función asociada a la prima de mercado

El comportamiento de esta función fue expresado a partir de una simulación de distribución normal según los datos en un rango para la prima de mercado entre -10 a 10, con una media de 0.55 y una desviación estándar de 0.38 siendo la prueba de bondad de ajuste de Kolmogórov-Smirnov con un p-valor de 85.57%, en donde el área sombreada resaltante muestra una concentración de valores en donde el valor del rendimiento es mayor al rendimiento de mercado, tal como se puede observar a partir de que la prima en relación al mercado se vuelve positiva los valores tienen a aumentar y aproximarse al valor de la unidad, mientras que si la diferencia es menor este se acercara a 0 por lo que ya dentro de la función se considera la restricción de castigar a la función si es que los valores son pequeños o negativos

$$\lambda_4 \left( \frac{1}{1 + e^{-(R_p - R_m)}} \right) \quad \text{Ec. 7}$$

Finalmente, el último factor a considerar es la diversificación del portafolio expresado como el promedio de la correlación de los activos que la conforman, y cuyo comportamiento está relacionado a una medida de penalización o premio a la función. La expresión matemática es la siguiente:

$$\lambda_5 \frac{2\rho_{ij}}{n(n-1)} \quad \text{Ec. 8}$$

Donde

$\rho_{ij}$ : es la correlacion entre el activo  $i$  y  $j$  que conforman el portafolio

$\frac{n(n-1)}{2}$ : Cantidad de combinaciones totales de activos que conforman el portafolio

considerando todos los factores antes mencionados se formula la siguiente función objetivo que las condensa:

$$\text{MAX (Objective Function)} = \frac{\lambda_1 R_P - \lambda_2 \sigma + \lambda_3 \left( \frac{R_p - R_f}{\sigma} \right) + \lambda_4 \left( \frac{1}{1 + e^{-(R_p - R_m)}} \right)}{\lambda_5 \left( \frac{2\rho_{ij}}{n(n-1)} \right)}$$

Ec. 9

Donde:

$\lambda_i$ : Son los pesos asignados a cada factor de acuerdo a su nivel de representatividad o importancia en relación a la maximización de la función objetivo, para  $i = 1,2,3,4,5$ , el valor para los pesos asignados en un proceso de selección del algoritmo genético se estableció de acuerdo al criterio del investigador, en relación a esto se hizo una simulación de Montecarlo haciendo uso de la herramienta Risk Simulator, para verificar la sensibilidad de estos factores con una cantidad de 10000 simulaciones. Para este caso en particular se usará el coeficiente encontrado como referencia para ponderar cada uno de los valores de lambda, que tal como se puede observar y en coherencia de la particularidad de cada factor asociado en donde los lambdas asociadas a la volatilidad y a la diversificación tienen valores negativos.

Se procedió a hacer la ponderación empírica de los lambdas a partir de una escala de 0 a 1, estos valores proporcionales a su nivel de correlación tal como se muestra la tabla 6:

Tabla 6. Proporción asignada de lambda's a cada factor del fitness

	<b>Correlación</b>	<b>% asignado</b>
Lambda 1	0.0096	1.0194%
Lambda 2	0.0076	0.8028%
Lambda 3	0.0159	1.6925%
Lambda 4	0.0016	0.1741%
Lambda 5	0.9069	96.3112%
Total	0.94	100.0000%

Considerando estos valores la ecuación tendrá la siguiente representación:

$$MAX(F. O) = \frac{0.0102 * R_p - 0.008 * \sigma + 0.017 * \left(\frac{R_p - R_f}{\sigma}\right) + 0.0017 * \left(\frac{1}{1 + e^{-(R_p - R_m)}}\right)}{0.963 * \left(\frac{2\rho_{ij}}{n(n-1)}\right)}$$

### **Hiperparámetros del algoritmo genético**

Determinar los hiperparámetros es una etapa fundamental en el desarrollo, diseño y evaluación del algoritmo genético. Para este caso específico, se utilizaron ciertas configuraciones basadas en el tipo de datos, el entorno del problema y la experiencia del investigador. Para el tamaño de la población inicial, el algoritmo se evaluó en tres espacios de búsqueda diferentes de 1000, 2500 y 5000 individuos, y el número de generaciones varía entre 50, 100 y 500, ajustado según la convergencia del algoritmo. Además, la probabilidad de cruce se establece en un 80%, considerando que los números de punto flotante del algoritmo son individuos y cubren un amplio espacio de búsqueda. En cuanto a la probabilidad de mutación, se elige un valor específico del 9% para este caso. En términos del método de selección, se utiliza el mecanismo de torneo, donde se eligen y evalúan un número específico de tres individuos. El operador de cruce utilizado es el CxBlend, conocido por generar nuevos individuos que exploran ampliamente los espacios de búsqueda al producir descendencia con características de ambos padres de manera lineal, ajustándose al tipo de problema. El operador de mutación utilizado es la mutación gaussiana con una media de 0 y una desviación estándar de 0.01, ya que se requiere generar individuos con un punto de mutación porcentual.

### **Evaluación del algoritmo genético**

El modelo procesado mediante el algoritmo 2 cuenta con la función de aptitud propuesta en este estudio y fue analizado utilizando un diseño experimental con un arreglo factorial de 2x3. Los factores considerados son el tamaño de la población inicial, denominado TPo, y el número de generaciones, denominado NG, cada uno con 3 niveles: 50, 100 y 500 para TPo, y 1000, 2500 y 5000 para NG, respectivamente. Para cada combinación factorial, se realizaron 4 repeticiones, como se muestra en la tabla 7:

---

**Algorithm 2** Algoritmo Genético
 

---

- 1: **Entrada:** Librerías.
- 2: **Entrada:** Base de datos de los activos del mercado.
- 3: **Asignación de hiperparámetros:** probabilidad de cruce, probabilidad de mutación  $\leftarrow 0.8, 0.09$ .
- 4: **Asignación:** Tasa libre de riesgo (Rf), Rendimiento del mercado Rm  $\leftarrow 0.002803, 0.06775$ .
- 5: **Asignación:** Población inicial, Número de generaciones  $\leftarrow 50, 2500$ .
- 6: **Asignación:**  $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \leftarrow 0.0096, 0.0076, 0.0159, 0.0016, 0.9069$
- 7: **Función** calcular rendimiento portafolio.  $\rightarrow R_p = \sum_{i=1}^n w_i \cdot R_i$
- 8: **Función** calcular volatilidad.  $\rightarrow \sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \sigma_i \cdot \sigma_j \cdot \rho_{ij}$
- 9: **Función** calcular correlación promedio.
- 10: **Función** calcular función objetivo.
- 11: **Función** crear individuo.
- 12: **Función** calcular estabilidad.
- 13: **Función** Algoritmo genético
- 14: Definición del problema (Maximización).
- 15: Evaluar a individuos

$$MAX(F.O) = \frac{\lambda_1 R_p - \lambda_2 \sigma + \lambda_3 \left( \frac{R_p - R_f}{\sigma} \right) + \lambda_4 \left( \frac{1}{1 + e^{-(R_p - R_M)}} \right)}{\lambda_5 \left( \frac{2\rho_{21}}{n(n-1)} \right)} \quad (5)$$

- 16: Realizar Cruce CxBlend.

$$D_i = P_{1i} + 0.8 \cdot (P_{2i} - P_{1i}) \quad (6)$$

- 17: Realizar mutación Gaussiana.

$$mutGaussian(x, \mu, \sigma, pb) = x + N(0, .01^2) \cdot 0.09 \quad (7)$$

- 18: Realizar selección por torneo.

$$selTournament_3(population, tournsize) = Mejor_1 \cup Mejor_2 \cup Mejor_3 \quad (8)$$

- 19: **Imprimir** el mejor fitness y el mejor individuo.
- 

Tabla 7. Repeticiones para cada configuración del algoritmo genético

NG	Tpo	Repeticiones			
		I	II	III	IV
50	1000	0.726487	0.752226	0.752226	0.752226
50	2500	0.772482	0.772482	0.772482	0.772482
50	5000	0.772482	0.772482	0.772482	0.772482
100	1000	0.752226	0.752226	0.752226	0.752226
100	2500	0.772482	0.772482	0.772482	0.772482
100	5000	0.772482	0.772482	0.772482	0.772482
500	1000	0.752226	0.752226	0.752226	0.752226
500	2500	0.772482	0.772482	0.772482	0.772482
500	5000	0.772482	0.772482	0.772482	0.772482

---

Después de haber generado las repeticiones se procedió a realizar en análisis de varianza con los siguientes resultados mostrados en la tabla 8:

Tabla 8. Análisis de varianza

Fuente de Variación	Gl	SC	CM	F	Sig.
NG	2	0.000037	0.000018	1	0.3811
TPo	2	0.004014	0.002007	109.0714	**0.0000
NG*TPo	4	0.000074	0.000018	1	0.4247
Error	27	0.000497	0.000018		
Total	35	0.004622			
CV	0.0024%				

Tal como se puede observar se encontró un nivel de significancia bilateral en el factor TPo, mientras que el factor NG, y la interacción entre los factores NGxTPo no tuvieron significancia estadística. También se puede evidenciar un coeficiente de variación de 0.0024% siendo un valor muy bajo lo que evidencia la estabilidad y la baja dispersión respecto a la media de los valores. En cuanto a la evaluación de pares de grupos, se realizó la prueba de Duncan para examinar más a fondo las diferencias entre grupos dentro del nivel del factor TPo. Esto se ilustra en la tabla 9:

Tabla 9. Prueba de Duncan

Contraste	A	B	dof	alternative	p-unc	BF10
					5.45E-	
Tpo	1000	2500	22	two-sided	10	1.087e+07
					5.45E-	
Tpo	1000	5000	22	two-sided	10	1.087e+07
Tpo	2500	5000	22	two-sided		nan

Según los hallazgos de la prueba, hay una diferencia significativa entre los niveles 1000 y

2500, mientras que no se encontró una diferencia significativa entre los niveles 2500 y 5000. Por lo tanto, la instancia del algoritmo genético se establece con un valor de NG de 50 y un TPo de 2500. El comportamiento con respecto a la estabilidad del algoritmo se puede observar en la figura 3.

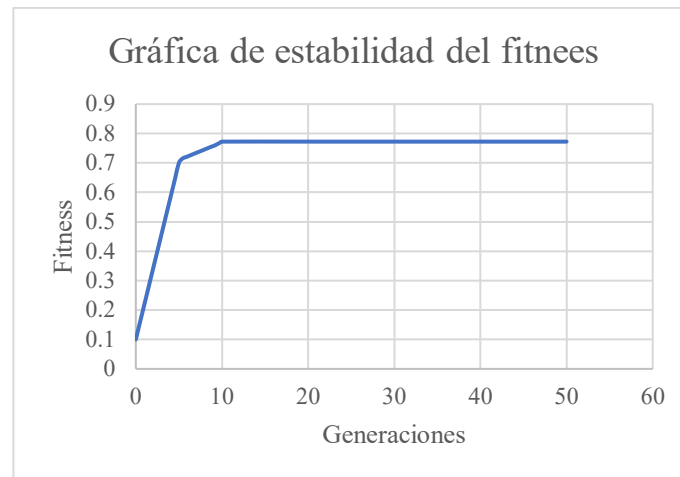


Figura 3. Representación de la estabilidad del algoritmo en función del fitness

Se puede observar en la figura 3 el fitness alcanza su convergencia a partir de la décima generación. Teniendo en cuenta que estas variables se oponen entre sí, con la rentabilidad siendo el elemento a maximizar y la volatilidad siendo el elemento a minimizar, se generó la frontera de Pareto para visualizar el comportamiento de estas variables a lo largo de las generaciones. Esto se muestra en la figura 4.

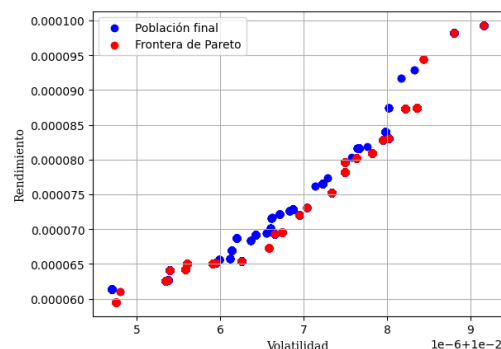


Figura 4. Frontera de Pareto

Como menciona Markowitz (1959), estos dos elementos son opuestos, y asumir una mayor rentabilidad implica asumir un mayor riesgo o volatilidad. Por lo tanto, la frontera de Pareto mostrada anteriormente evidencia este comportamiento, lo cual se alinea con la naturaleza de estas variables, y no se observa ninguna anomalía en su relación lineal. Finalmente, al conocer la estabilidad del algoritmo, podemos representar al individuo que logra la optimalidad en este punto, como se muestra en la siguiente tabla 10.

Tabla 10. Porcentajes de capital a invertir en los activos

<b>% a invertir</b>	<b>Empresas</b>
0.2948%	CASAGRC1
6.1901%	BUENAVC1
2.5055%	BBVAC1
2.2108%	ATCU
7.0744%	FALABEC1
4.1268%	ELCOMEI1
1.0317%	CVERDEC1
4.6426%	CREDITC1
0.1474%	CPACASC1
6.4849%	CORAREC1
4.2741%	LUISAI1
5.0847%	INVCENC1
6.9270%	INTERBC1
2.8740%	IFS
4.4952%	SCOTIAC1
2.2845%	PML
1.7686%	MINSURI1
3.2424%	MIRL
2.5792%	NEXAPEC1
5.1584%	PERUBAI1
6.6323%	POSITIC1



6.1164%	TELEFBC1
7.0007%	TEF
6.8534%	SAGAC1

---

Tal como se puede observar las empresas con el mayor porcentaje a invertir fueron las representadas por FALABEC1 y TEF, con un 7% aproximadamente, mientras que las empresas que contempla el menor porcentaje a invertir fue CPACASC1 con un 0.1474%

## DISCUSIÓN

Para la discusión de la presente investigación con diversos estudios, se debe considerar que cada autor dentro de los antecedentes y la presente investigación ha desarrollado su estudio en un entorno particular, es decir los resultados obtenidos referente a la rentabilidad o volatilidad depende específicamente de los activos de capital escogidos y evaluados por los autores.

Maholi et al. (2019) de forma similar a esta investigación usa una metodología en donde primero se usa una red neuronal artificial. En el caso del autor la arquitectura usada es de una red densa que sirve en la fase de selección de activos de capital más aptos para conformar los portafolios generados por el algoritmo genético, el cual tendrá como función de aptitud el retorno por sobre el riesgo, el cual denomina ERB, que un símil del radio de Sharpe y que mide el exceso de rentabilidad frente a una tasa base. La principal crítica respecto a esta metodología es que las redes neuronales densas no disciernen en cuanto a las series de tiempo, es decir las entradas no son diferenciadas por lo que los datos se adaptan más a la estructura interna que tienen, en vez de dar una respuesta en relación a la variable tiempo en donde la precisión alcanza un 98% para un modelo unistep, el cual es inferior a esta investigación en donde este modelo tiene una precisión del 99% en una red LSTM, mientras que el MAE obtenido por el investigador es de 5.60% mientras que para la presente investigación fue de 5.07% en modelo multistep. Entre los resultados respecto al retorno del portafolio, este alcanzo un 1.42% sin embargo la volatilidad del portafolio es mucho mayor en puntos porcentuales alcanzando un 1.86% siendo muy superior al 0.00612188% encontrado en el portafolio óptimo de esta investigación.

Rodríguez et al. (2020) en su investigación utiliza otro tipo de activos los cuales son las monedas de diversos países, para minimizar el riesgo debido a la estabilidad que estas representan, estas

fueron codificadas en cuanto a su rentabilidad usando logaritmos neperianos. Para el caso se los operadores genéticos utilizan en forma análoga el cblend y la mutación gaussiana, sin embargo, la función objetivo está relacionada a la convencional y usada en muchas fuentes, la cual es la asociada a la rentabilidad por cada grado de riesgo asumido. Para el caso de esta investigación difiere en el resultado de la ponderación. Para el caso de los resultados cuantificados, se obtuvo un 0.049292% de rendimiento, valor mucho menor a la propuesta de esta investigación mientras que el riesgo o volatilidad fue de 0.0354% mucho mayor que la volatilidad encontrada en la presente investigación.

Bo Liu (2023) en su investigación genera un enfoque de red para optimizar el portafolio de inversión, en donde la entropía de la red se usa como medida para conocer el grado de relevancia de cada acción que conforma el portafolio que pasara a ser optimizado por un algoritmo genético, esto es un análogo del preprocesamiento de esta investigación, en donde se usa una red neuronal para seleccionar las acciones con mejor desempeño en sus próximos valores para pasar a la optimización del algoritmo genético, tal como se mencionó la principal métrica de la investigación es la entropía de la red que representa el portafolio de inversión, mientras que en la presente investigación se usó un método más relacionado al comportamiento sistemático del mercado. Los resultados encontrados con esta presente investigación difieren en cuanto a la rentabilidad siendo el resultado mayor, ya que para Bo Liu (2023) solo encontró un rendimiento de 0.2667% de retorno. Por otra parte, Chun-Hao et al. (2019) en su investigación en contraproposición a esta investigación es la consideración de múltiples aspectos en la función objetivo en donde se tiene en cuenta que el portafolio este balanceado y la satisfacción relacionado al portafolio, así como métricas de inversión teniendo finalmente una fitness de 4 elementos en su numerador y un elemento en su denominador, para el caso de esta investigación otro aspecto que difiere es el número de elementos que conforman la cartera, en donde solo se evaluó 5 activos de capital, basándose principalmente en el rendimiento sobre la inversión obteniendo como resultados un máximo de 0.6% de retorno siendo inferior al rendimiento obtenido a los resultados encontrados en esta investigación.

Vasiani (2020) en su investigación considera un índice de prioridad por cada sector empresarial a evaluar por lo que se considera un preprocesamiento que lleva a los activos con mayor puntuación por cada sector para luego pasar a la optimización del portafolio de inversión, este proceso difiere

al de la presente investigación teniendo en cuenta que no toma series de tiempo sino los valores de comportamiento interno de la empresa en cuanto a sus transacciones, como la compra, venta de activos o devaluación de los mismo, otro aspecto resaltante es la función objetivo en donde se usó el retorno por cada unidad de riesgo, para este caso tuvo un promedio de 3% superior al de esta investigación, sin embargo no se hace referencia al portafolio de inversión en relación a la volatilidad de los activos por lo que esta rentabilidad podría estar relacionada a un riesgo muy alto de inversión.

## 5. Conclusiones

Se desarrolló una red neuronal artificial para predecir las tendencias en activos financieros, configurada con 150 unidades de procesamiento, entrenada durante 400 épocas y ajustada con una tasa de aprendizaje de 0.0004. Los resultados demostraron una precisión significativa del 92.35%, posicionándola como una herramienta confiable para la predicción en este contexto. Además, se evaluaron métricas de error, como el error cuadrático medio cuyo valor fue de 6.33% y un error absoluto medio cuyo valor fue de 5.07%, lo que indica una buena capacidad de generalización del modelo.

En cuanto al algoritmo genético, la función objetivo se diseñó considerando múltiples aspectos clave como la rentabilidad, volatilidad, coeficiente de Sharpe, prima de riesgo de mercado y diversificación de la cartera. Esta modelización es crucial para comprender y optimizar la asignación de activos dentro de la cartera de inversiones, maximizando los rendimientos esperados mientras simultáneamente se minimiza el riesgo asociado.

La representación cromosómica de los individuos en el algoritmo genético se basó en un vector de porcentajes de capital, lo que permitió la manipulación efectiva de la asignación de activos dentro de la cartera. Además, se ajustaron los hiperparámetros del algoritmo como el tamaño de la población inicial, el número de generaciones, la probabilidad de cruce y la probabilidad de mutación para lograr un rendimiento óptimo del algoritmo.

Después de numerosas iteraciones y evaluaciones, se determinó que la mejor configuración del algoritmo genético fue aquella que comprendía 50 generaciones e individuos de población inicial de 2500. Esta configuración logró un fitness de 0.772482, un rendimiento del 1.00058% y

una volatilidad del 0.00612%, lo que sugiere una cartera bien equilibrada con un alto rendimiento y bajo riesgo.

La metodología desarrollada en esta investigación ha demostrado resultados favorables, por lo que puede ser aplicada en cualquier Bolsa de Valores, siempre y cuando exista activos de renta variable.

## REFERENCIAS

- [1] Ames Santillán, J. C. (2012). Alternativas de diversificación internacional para portafolios de acciones de la Bolsa de Valores de Lima. 7, 13. <https://www.redalyc.org/articulo.oa?id=281623577003>
- [2] Bennett, D. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25(5), 464–469. <https://doi.org/https://doi.org/10.1111/j.1467-842X.2001.tb00294.x>
- [3] Bo, L. (2023). Portfolio Optimization Based on Complex Networks and Genetic Algorithms. *Academic Journal of Computing & Information Science*, 6(10). <https://doi.org/10.25236/ajcis.2023.061013>
- [4] Cáceres Chian, V. A. E. (2018). Predicción de precios de acciones de bolsa de valores utilizando support vector regression.
- [5] Candia Garcia, C. D., López Castro, L. F., & Jaimes Suárez, S. A. (2020). Selección óptima del portafolio de proyectos utilizando metaheurísticas de población y trayectoria meta-optimizadas. *Revista EIA*, 17(34), 1–18. <https://doi.org/10.24050/reia.v17i34.1399>
- [6] Dagnino, J. (2014). Datos faltantes (Missing values). *Revista chilena de anestesia*, 43, 332–334. <https://revistachilenadeanestesia.cl/datos-faltantes-missing-values/>
- [7] Das, A., Chaudhuri, T., Roy, S. S., Biswas, S., & Guha, B. (2023). Selection of Appropriate Portfolio Optimization Strategy. *Theoretical and Applied Computational Intelligence*, 1(1), 58–81. <https://doi.org/10.31181/taci1120237>
- [8] Didier, T., Levine, R., Llovet Montanes, R., & Schmukler, S. L. (2021). Capital market financing and firm growth. *Journal of International Money and Finance*, 118. <https://doi.org/10.1016/j.jimonfin.2021.102459>

- [9] Fang, C., & Wang, C. (2020). Time Series Data Imputation: A Survey on Deep Learning Approaches. <http://arxiv.org/abs/2011.11347>
- [10] Gitman, L. J., & Joehnk, M. D. (2009). *Fundamentos de inversiones*. Pearson Educación.
- [11] Gutiérrez Reina, D., Tapia Córdoba, A., & del Nozal, Á. (2020). *Algoritmos Genéticos con Python (Primera)*. Alfaomega-Marcombo.
- [12] Hart, O. D., & Zingales, L. (2011). Inefficient provision of liquidity. NBER working paper series. <http://www.nber.org/papers/w17299>
- [13] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1), 388–427. <https://doi.org/10.1016/j.ijforecast.2020.06.008>
- [14] Li, C., Shen, L., & Qian, G. (2023). Online Hybrid Neural Network for Stock Price Prediction: A Case Study of High-Frequency Stock Trading in the Chinese Market. *Econometrics*, 11(2). <https://doi.org/10.3390/econometrics11020013>
- [15] Li, Q., Al-Jaifi, H. A., & Kamaruddin, N. (2023). Can Long-short Term Memory Neural Network with Symbolic Genetic Algorithm Predict Stock Price Change Basing on Fundamental Indicators. <https://doi.org/10.21203/rs.3.rs-3306087/v1>
- [16] Liagkouras, K. (2019). A new three-dimensional encoding multiobjective evolutionary algorithm with application to the portfolio optimization problem. *Knowledge-Based Systems*, 163, 186–203. <https://doi.org/10.1016/j.knosys.2018.08.025>
- [17] Lim, S., Kim, M. J., & Ahn, C. W. (2020). A Genetic Algorithm (GA) Approach to the Portfolio Design Based on Market Movements and Asset Valuations. *IEEE Access*, 8, 140234–140249. <https://doi.org/10.1109/ACCESS.2020.3013097>
- [18] Maholi Solin, M., Alamsyah, A., Rikumahu, B., & Arya Saputra, M. A. (2019). Forecasting Portfolio Optimization using Artificial Neural Network and Genetic Algorithm. *IEEE Xplore*, 1–7. <https://doi.org/10.1109/ICoICT.2019.8835344>
- [19] Maholi Solin, M., Alamsyah, A., Rikumahu, B., & Arya Saputra, M. A. (2019). Forecasting Portfolio Optimization using Artificial Neural Network and Genetic Algorithm. *IEEE Xplore*, 1–7. <https://doi.org/10.1109/ICoICT.2019.8835344>

- [20] Markowitz, H. (1952). Portfolio Selection. *Journal of Economic Literature*, 7(1), 77–91. <http://links.jstor.org/sici?sici=0022-1082%28195203%297%3A1%3C77%3APS%3E2.0.CO%3B2-1>
- [21] Rodríguez López, E., García Mejía, J. F., & Martínez Garduño, Y. (2020). Diseño de un portafolio de inversión con algoritmos evolutivos. *Saeta Digital*, 6(1), 51–70.
- [22] Ruochen, L., & Muchao, L. (2021). Stock Trend Prediction Algorithm Based on Deep Recurrent Neural Network. *Wireless Communications and Mobile Computing*, 2021. <https://doi.org/10.1155/2021/5694975>
- [23] Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series USA. [https://doi.org/978-1-7281-0858-2/19/\\$31.00](https://doi.org/978-1-7281-0858-2/19/$31.00)
- [24] Vasiani, V. D., Handari, B. D., & Hertono, G. F. (2020). Stock portfolio optimization using priority index and genetic algorithm. *Journal of Physics: Conference Series*, 1442(1). <https://doi.org/10.1088/1742-6596/1442/1/012031>