

Sistema Híbrido para Clasificación de Perfiles de Riesgo y Predicción de Series Temporales Financieras mediante Redes Neuronales Competitivas, Mapas Auto-Organizados y Optimización Metaheurística

Un Enfoque Integrado de Machine Learning e Ingeniería Financiera

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE
DEPARTAMENTO DE MATEMÁTICAS Y FÍSICA
MÉTODOS DE INTELIGENCIA ARTIFICIAL

Profesor: César Lozano Díaz

Integrantes:

Santiago Ortiz García
Paula Inés Pelayo Morales
Roberto Carlos Guzmán Orduño
Christopher Jacob Ahumada Robles

Diciembre 2025

December 4, 2025

Abstract

Este documento presenta un marco metodológico integral para la clasificación de activos financieros según su perfil de riesgo y la predicción de sus precios mediante técnicas avanzadas de aprendizaje automático. El sistema propuesto integra múltiples paradigmas de clustering no supervisado—incluyendo redes neuronales competitivas, mapas auto-organizados de Kohonen (SOM), K-Means, modelos de mezclas gaussianas (GMM), clustering jerárquico aglomerativo y Fuzzy C-Means—con estrategias de optimización metaheurística basadas en Particle Swarm Optimization (PSO) y algoritmos genéticos (GA) para la construcción de ensamblajes óptimos. La componente predictiva emplea arquitecturas neuronales de complejidad creciente: ADALINE con regularización L_2 , perceptrones multicapa (MLP) con retropropagación Levenberg-Marquardt, y redes LSTM para captura de dependencias temporales. El framework se complementa con optimización de portafolios mediante simulación Monte Carlo y PSO, siguiendo los principios de la teoría moderna de portafolios de Markowitz. Los resultados experimentales demuestran que el modelo híbrido RedCompPSO alcanza un F1-Score de 0.9166, superando significativamente a los métodos individuales.

Contents

| | | |
|----------|--|----------|
| 1 | Introducción y Motivación Teórica | 3 |
| 1.1 | Contexto del Problema | 3 |
| 1.2 | Objetivos del Sistema | 3 |
| 2 | Arquitectura del Sistema y Pipeline Computacional | 3 |
| 2.1 | Diagrama de Flujo General | 3 |
| 2.2 | Formalización del Espacio de Características | 4 |
| 2.2.1 | Volatilidad Anualizada | 4 |
| 2.2.2 | Retorno Anualizado | 4 |
| 2.2.3 | Ratio de Sharpe | 4 |

| | | |
|----------|---|-----------|
| 2.2.4 | Maximum Drawdown | 4 |
| 2.2.5 | Value at Risk Paramétrico | 5 |
| 3 | Construcción del Ground Truth mediante Score Compuesto | 5 |
| 3.1 | Normalización Min-Max | 5 |
| 3.2 | Score de Riesgo Compuesto | 5 |
| 3.3 | Discretización en Categorías | 5 |
| 4 | Algoritmos de Clustering: Formulación Matemática | 6 |
| 4.1 | Red Neuronal Competitiva | 6 |
| 4.2 | Mapa Auto-Organizado de Kohonen (SOM) | 6 |
| 4.3 | K-Means | 7 |
| 4.4 | Modelo de Mezclas Gaussianas (GMM) | 7 |
| 4.5 | Clustering Jerárquico Aglomerativo (Ward) | 7 |
| 4.6 | Fuzzy C-Means (FCM) | 7 |
| 5 | Optimización Metaheurística para Clustering | 7 |
| 5.1 | Particle Swarm Optimization (PSO) | 7 |
| 5.2 | Algoritmo Genético para Ensemble | 7 |
| 6 | Modelos Predictivos de Series Temporales | 8 |
| 6.1 | ADALINE con Regularización L_2 | 8 |
| 6.2 | Perceptrón Multicapa (MLP) | 8 |
| 6.3 | Long Short-Term Memory (LSTM) | 9 |
| 7 | Optimización de Portafolio | 9 |
| 7.1 | Teoría Moderna de Portafolios | 9 |
| 8 | Resultados Experimentales | 9 |
| 8.1 | Resultados de Clustering | 10 |
| 8.1.1 | Comparación de Modelos de Clustering | 10 |
| 8.1.2 | Exactitud vs Coeficiente Silhouette | 10 |
| 8.1.3 | Métricas por Clase del Mejor Modelo | 10 |
| 8.1.4 | Visualización de Pesos Aprendidos | 12 |
| 8.1.5 | Clasificación en el Espacio Riesgo-Retorno | 12 |
| 8.2 | Selección de Activos y Configuración del Portafolio | 12 |
| 8.3 | Resultados de Optimización de Portafolio | 13 |
| 8.4 | Resultados de Predicción de Precios | 14 |
| 8.4.1 | Convergencia del Entrenamiento ADALINE | 14 |
| 8.4.2 | Comparación de Modelos Predictivos | 14 |
| 8.5 | Resumen de Resultados | 16 |
| 9 | Conclusiones | 16 |
| 9.1 | Contribuciones Principales | 16 |
| 9.2 | Limitaciones y Trabajo Futuro | 16 |
| A | Resumen de Hiperparámetros | 17 |

1 Introducción y Motivación Teórica

1.1 Contexto del Problema

La gestión cuantitativa de portafolios enfrenta el desafío fundamental de caracterizar activos financieros según su comportamiento riesgo-retorno de manera que permita decisiones de inversión informadas. Sea $\mathcal{A} = \{a_1, a_2, \dots, a_N\}$ un universo de N activos cotizados, donde cada activo a_i genera una serie temporal de precios $\{P_i(t)\}_{t=1}^T$.

Definición 1.1 (Rendimiento Logarítmico). *Para un activo a_i con precio $P_i(t)$ en el instante t , el rendimiento logarítmico se define como:*

$$r_i(t) = \ln \left(\frac{P_i(t)}{P_i(t-1)} \right) = \ln P_i(t) - \ln P_i(t-1) \quad (1)$$

La elección del rendimiento logarítmico sobre el rendimiento simple $R_i(t) = \frac{P_i(t) - P_i(t-1)}{P_i(t-1)}$ se justifica por su propiedad de aditividad temporal:

$$r_i(t_1, t_n) = \sum_{k=1}^{n-1} r_i(t_k, t_{k+1}) \quad (2)$$

1.2 Objetivos del Sistema

El sistema desarrollado aborda tres problemas interrelacionados:

- P1. Clasificación de Perfiles de Riesgo:** Dado un conjunto de métricas $\mathbf{x}_i \in \mathbb{R}^d$ para cada activo, encontrar una partición $\mathcal{C} = \{C_1, C_2, C_3, C_4\}$ que agrupe activos con perfiles similares.
- P2. Predicción de Precios:** Para activos seleccionados, construir un modelo predictivo $\hat{P}(t+1) = f(\mathbf{P}_{t-w+1:t})$ donde w es la ventana temporal.
- P3. Optimización de Portafolio:** Determinar pesos óptimos $\mathbf{w}^* \in \Delta^{n-1}$ que maximicen el ratio de Sharpe.

2 Arquitectura del Sistema y Pipeline Computacional

2.1 Diagrama de Flujo General

El sistema opera en cuatro fases secuenciales con retroalimentación entre componentes:

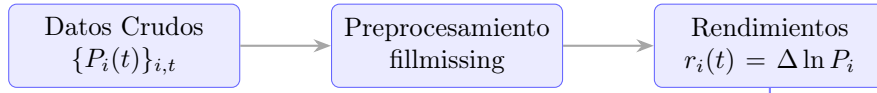
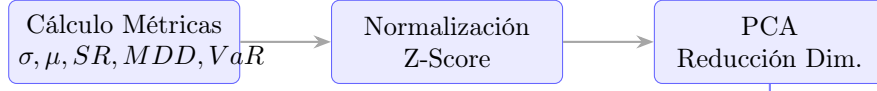
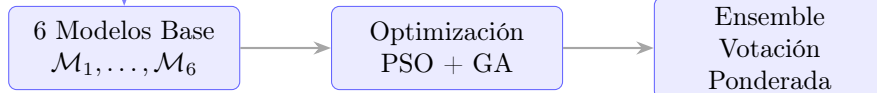
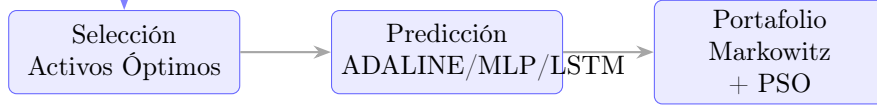
FASE 1: Ingesta de Datos**FASE 2: Ingeniería de Características****FASE 3: Clustering Híbrido****FASE 4: Predicción y Portafolio**

Figure 1: Pipeline computacional del sistema de análisis financiero organizado en 4 fases.

2.2 Formalización del Espacio de Características

Definición 2.1 (Espacio de Métricas de Riesgo). Para cada activo a_i , definimos el vector de características $\mathbf{x}_i \in \mathbb{R}^5$ como:

$$\mathbf{x}_i = \begin{pmatrix} \sigma_i^{(ann)} \\ MDD_i \\ VaR_{i,0.95} \\ \kappa_i \\ -SR_i \end{pmatrix} \quad (3)$$

donde cada componente captura un aspecto diferente del riesgo.

Las métricas se definen formalmente como sigue:

2.2.1 Volatilidad Anualizada

$$\sigma_i^{(ann)} = \sqrt{252} \cdot \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_i(t) - \bar{r}_i)^2} \quad (4)$$

2.2.2 Retorno Anualizado

$$\mu_i^{(ann)} = 252 \cdot \bar{r}_i = 252 \cdot \frac{1}{T} \sum_{t=1}^T r_i(t) \quad (5)$$

2.2.3 Ratio de Sharpe

$$SR_i = \frac{\mu_i^{(ann)}}{\sigma_i^{(ann)}} = \frac{\mathbb{E}[r_i] \cdot 252}{\sqrt{\text{Var}[r_i] \cdot 252}} \quad (6)$$

2.2.4 Maximum Drawdown

$$MDD_i = \max_{t \in [1, T]} \left\{ \frac{\max_{\tau \leq t} P_i(\tau) - P_i(t)}{\max_{\tau \leq t} P_i(\tau)} \right\} \quad (7)$$

2.2.5 Value at Risk Paramétrico

Asumiendo normalidad de rendimientos:

$$VaR_{i,\alpha} = -(\bar{r}_i + z_\alpha \cdot \sigma_i) \quad (8)$$

donde $z_\alpha = \Phi^{-1}(\alpha)$ es el cuantil de la distribución normal estándar. Para $\alpha = 0.05$, $z_{0.05} \approx -1.645$.

3 Construcción del Ground Truth mediante Score Compuesto

3.1 Normalización Min-Max

Para homogeneizar las escalas de las métricas, aplicamos normalización min-max:

$$\tilde{x}_i^{(j)} = \frac{x_i^{(j)} - \min_k x_k^{(j)}}{\max_k x_k^{(j)} - \min_k x_k^{(j)} + \epsilon} \quad (9)$$

3.2 Score de Riesgo Compuesto

Definición 3.1 (Score de Riesgo Ponderado). *El score de riesgo para el activo i se define como la combinación convexa:*

$$S_i = \sum_{j=1}^4 w_j \cdot \tilde{x}_i^{(j)} \quad (10)$$

con pesos $\mathbf{w} = (0.35, 0.25, 0.25, 0.15)^\top$ correspondientes a (Volatilidad, MaxDD, VaR₉₅, Sharpe invertido).

3.3 Discretización en Categorías

Las etiquetas de ground truth $y_i \in \{1, 2, 3, 4\}$ se asignan mediante cuartiles:

$$y_i = \begin{cases} 1 \text{ (Conservador)} & \text{si } S_i \leq Q_1(S) \\ 2 \text{ (Moderado)} & \text{si } Q_1(S) < S_i \leq Q_2(S) \\ 3 \text{ (Arriesgado)} & \text{si } Q_2(S) < S_i \leq Q_3(S) \\ 4 \text{ (Muy Arriesgado)} & \text{si } S_i > Q_3(S) \end{cases} \quad (11)$$

4 Algoritmos de Clustering: Formulación Matemática

4.1 Red Neuronal Competitiva

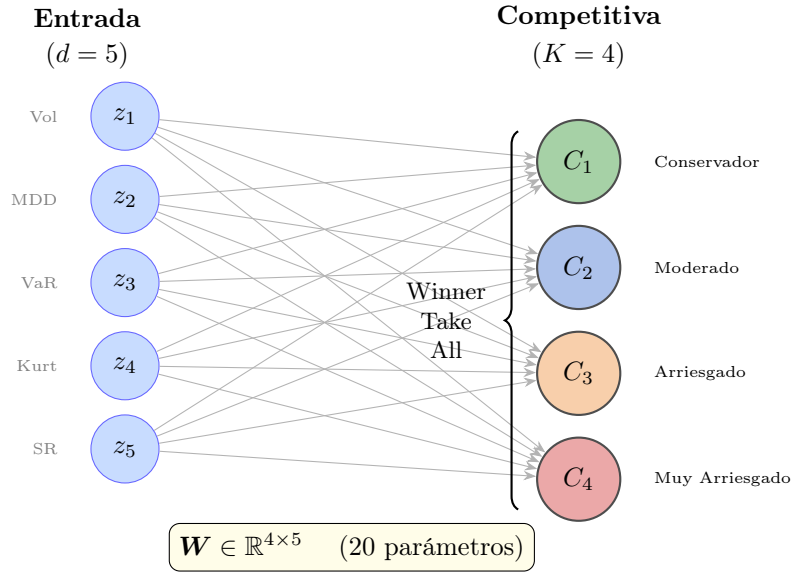


Figure 2: Arquitectura de la Red Neuronal Competitiva [5] → [4].

Para un patrón de entrada $\mathbf{z} \in \mathbb{R}^d$, el proceso de competición determina:

$$j^* = \arg \min_{j \in \{1, \dots, K\}} \|\mathbf{z} - \mathbf{w}_j\|^2 \quad (12)$$

La actualización de pesos sigue la regla de Kohonen:

$$\mathbf{w}_{j^*}^{(t+1)} = \mathbf{w}_{j^*}^{(t)} + \eta(t) \cdot (\mathbf{z} - \mathbf{w}_{j^*}^{(t)}) \quad (13)$$

4.2 Mapa Auto-Organizado de Kohonen (SOM)

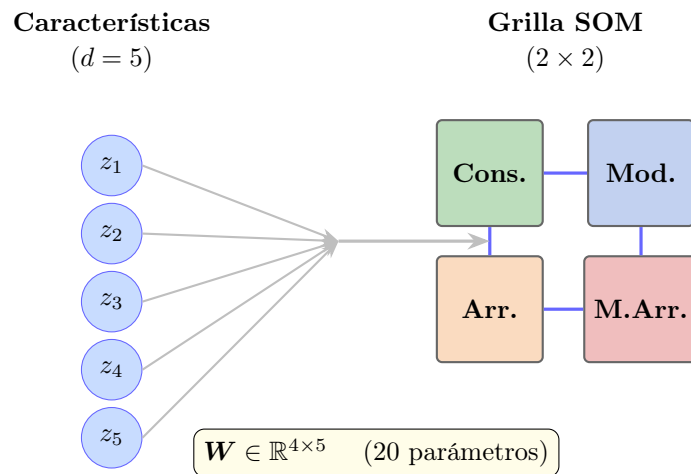


Figure 3: Mapa Auto-Organizado (SOM) con topología 2×2 .

La función de vecindad gaussiana se define como:

$$h(j, j^*, t) = \exp \left(-\frac{\|\mathbf{r}_j - \mathbf{r}_{j^*}\|^2}{2\sigma(t)^2} \right) \quad (14)$$

4.3 K-Means

Definición 4.1 (Problema de K-Means). *Encontrar centroides $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ y asignaciones $\{a_i\}_{i=1}^N$ que minimicen:*

$$J_{KM} = \sum_{i=1}^N \|\mathbf{z}_i - \mathbf{c}_{a_i}\|^2 = \sum_{k=1}^K \sum_{i:a_i=k} \|\mathbf{z}_i - \mathbf{c}_k\|^2 \quad (15)$$

4.4 Modelo de Mezclas Gaussianas (GMM)

Definición 4.2 (GMM). *Un GMM modela la densidad de datos como:*

$$p(\mathbf{z}) = \sum_{k=1}^K \pi_k \cdot \mathcal{N}(\mathbf{z} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (16)$$

4.5 Clustering Jerárquico Aglomerativo (Ward)

Definición 4.3 (Criterio de Ward).

$$\Delta(C_i, C_j) = \frac{|C_i| \cdot |C_j|}{|C_i| + |C_j|} \|\bar{\mathbf{z}}_i - \bar{\mathbf{z}}_j\|^2 \quad (17)$$

4.6 Fuzzy C-Means (FCM)

Definición 4.4 (FCM). *FCM minimiza la función objetivo:*

$$J_{FCM} = \sum_{i=1}^N \sum_{k=1}^K u_{ik}^m \|\mathbf{z}_i - \mathbf{c}_k\|^2 \quad (18)$$

sujeto a $\sum_{k=1}^K u_{ik} = 1$ y $u_{ik} \geq 0$, donde $m > 1$ es el exponente de fuzzificación.

5 Optimización Metaheurística para Clustering

5.1 Particle Swarm Optimization (PSO)

La dinámica del enjambre sigue:

$$\mathbf{v}_i^{(t+1)} = \omega(t) \mathbf{v}_i^{(t)} + c_1 \mathbf{r}_1 \odot (\mathbf{p}_i - \mathbf{x}_i^{(t)}) + c_2 \mathbf{r}_2 \odot (\mathbf{g} - \mathbf{x}_i^{(t)}) \quad (19)$$

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \mathbf{v}_i^{(t+1)} \quad (20)$$

La función fitness para optimizar la red competitiva es:

$$f_{PSO-Comp}(\mathbf{w}) = \text{Accuracy}(\text{CompNet}(\mathbf{Z} \odot \mathbf{w}^\top), \mathbf{y}) \quad (21)$$

5.2 Algoritmo Genético para Ensemble

Cada individuo codifica los pesos del ensemble:

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_6) \in \mathbb{R}_+^6 \quad (22)$$

La predicción del ensemble es:

$$\hat{\mathbf{y}}_i^{ens} = \arg \max_{k \in \{1, \dots, K\}} \sum_{m=1}^6 \alpha_m \cdot \mathbb{I}[\hat{\mathbf{y}}_i^{(m)} = k] \quad (23)$$

Función Fitness:

$$f_{GA-Ens}(\boldsymbol{\alpha}) = \text{Accuracy}(\hat{\mathbf{y}}^{ens}, \mathbf{y}) + 0.05 \cdot \bar{S}(\mathbf{Z}, \hat{\mathbf{y}}^{ens}) \quad (24)$$

6 Modelos Predictivos de Series Temporales

6.1 ADALINE con Regularización L_2

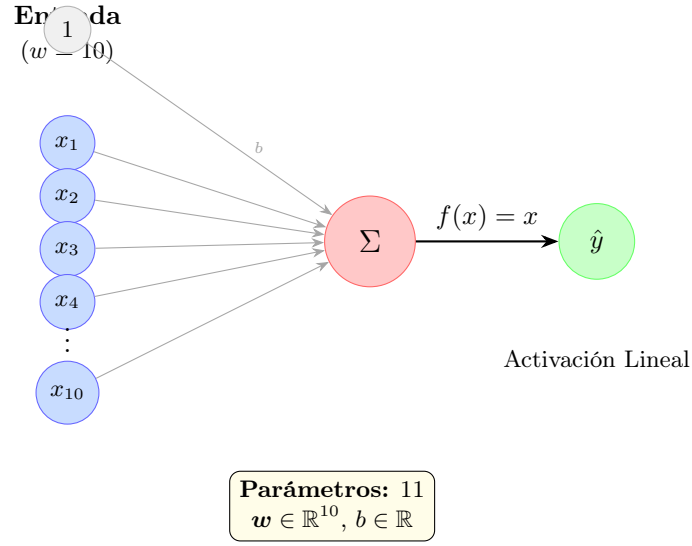


Figure 4: Arquitectura ADALINE: $[10] \rightarrow [1]$ con 11 parámetros.

Definición 6.1 (Función de Costo J). *La función de costo con regularización L_2 es:*

$$J(\mathbf{w}, b) = \frac{1}{2N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2 \quad (25)$$

donde $\hat{y}_i = \mathbf{x}_i^\top \mathbf{w} + b$ y $\lambda = 0.01$.

6.2 Perceptrón Multicapa (MLP)

Arquitectura: $[10] \rightarrow [16] \rightarrow [8] \rightarrow [1]$ con 321 parámetros.

$$\mathbf{h}^{(1)} = \tanh(\mathbf{W}^{(1)} \mathbf{x} + \mathbf{b}^{(1)}) \in \mathbb{R}^{16} \quad (26)$$

$$\mathbf{h}^{(2)} = \tanh(\mathbf{W}^{(2)} \mathbf{h}^{(1)} + \mathbf{b}^{(2)}) \in \mathbb{R}^8 \quad (27)$$

$$\hat{y} = \mathbf{w}^{(3)\top} \mathbf{h}^{(2)} + b^{(3)} \in \mathbb{R} \quad (28)$$

6.3 Long Short-Term Memory (LSTM)

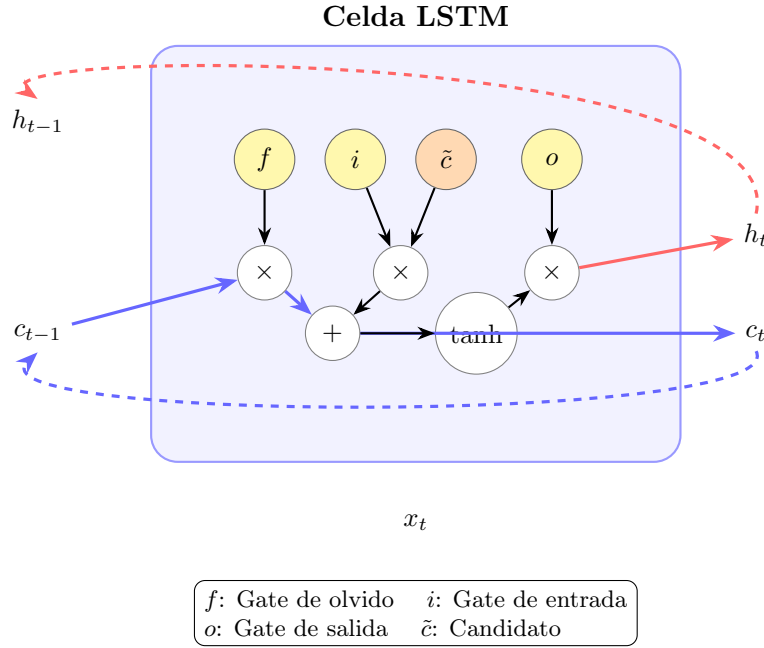


Figure 5: Celda LSTM con gates de olvido (f), entrada (i), y salida (o).

Ecuaciones de la celda LSTM:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \quad (29)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad (30)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) \quad (31)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \quad (32)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \quad (33)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (34)$$

7 Optimización de Portafolio

7.1 Teoría Moderna de Portafolios

$$\mathbb{E}[r_p] = \mathbf{w}^\top \boldsymbol{\mu} \quad (35)$$

$$\text{Var}[r_p] = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \quad (36)$$

$$SR_p = \frac{\mathbf{w}^\top \boldsymbol{\mu}}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}} \quad (37)$$

La optimización mediante PSO busca:

$$\mathbf{w}^* = \arg \max_{\mathbf{w} \in \Delta^{n-1}} SR(\mathbf{w}) \quad (38)$$

8 Resultados Experimentales

Esta sección presenta los resultados obtenidos en la aplicación del sistema híbrido al análisis de activos del Índice de Precios y Cotizaciones (IPC) de la Bolsa Mexicana de Valores.

8.1 Resultados de Clustering

8.1.1 Comparación de Modelos de Clustering

La Tabla 1 presenta las métricas de clasificación para los 9 modelos evaluados, incluyendo los 6 métodos base y los 3 enfoques híbridos optimizados.

Table 1: Comparación de métricas de clasificación para modelos de clustering.

| Modelo | Accuracy | Precision | Recall | F1-Score |
|-------------------|---------------|---------------|---------------|---------------|
| RedCompetitiva | 83.33% | 83.54% | 83.33% | 0.8331 |
| SOM | 58.33% | 48.33% | 58.33% | 0.4970 |
| KMeans | 55.56% | 46.25% | 55.56% | 0.4629 |
| GMM | 52.78% | 44.41% | 52.78% | 0.4663 |
| Jerárquico | 47.22% | 42.73% | 47.22% | 0.4057 |
| FCM | 55.56% | 50.09% | 55.56% | 0.5180 |
| EnsemblePSO | 83.33% | 83.54% | 83.33% | 0.8331 |
| RedCompPSO | 91.67% | 91.94% | 91.67% | 0.9166 |
| EnsembleGA | 83.33% | 83.54% | 83.33% | 0.8331 |

El modelo **RedCompPSO** (Red Competitiva optimizada con PSO) alcanza el mejor desempeño con un F1-Score de **0.9166**, superando significativamente a todos los demás métodos. La optimización PSO mejora el F1-Score de la red competitiva base de 0.8331 a 0.9166, representando una mejora del **10.02%**.

8.1.2 Exactitud vs Coeficiente Silhouette

La Tabla 2 muestra la relación entre la exactitud de clasificación y la calidad intrínseca del clustering medida por el coeficiente Silhouette.

Table 2: Exactitud y coeficiente Silhouette por modelo.

| Modelo | Exactitud | Silhouette |
|-------------------|---------------|---------------|
| RedCompetitiva | 0.8333 | 0.2427 |
| SOM | 0.5833 | 0.5701 |
| KMeans | 0.5556 | 0.5817 |
| GMM | 0.5278 | 0.2865 |
| Jerárquico | 0.4722 | 0.5326 |
| FCM | 0.5556 | 0.3313 |
| EnsemblePSO | 0.8333 | 0.2427 |
| RedCompPSO | 0.9167 | 0.2186 |
| EnsembleGA | 0.8333 | 0.2427 |

Es notable observar que existe una correlación negativa entre exactitud y Silhouette. Los métodos con mayor Silhouette (KMeans, SOM, Jerárquico) tienen menor exactitud, lo cual indica que la estructura geométrica óptima de los clusters no coincide necesariamente con la clasificación supervisada basada en el score de riesgo compuesto.

8.1.3 Métricas por Clase del Mejor Modelo

La Tabla 3 detalla el desempeño del modelo RedCompPSO para cada categoría de riesgo.

Table 3: Métricas por clase para RedCompPSO (mejor modelo).

| Categoría | Precision | Recall | F1-Score |
|----------------|-----------|---------|----------|
| Conservador | 90.00% | 100.00% | 0.9474 |
| Moderado | 88.89% | 88.89% | 0.8889 |
| Arriesgado | 88.89% | 88.89% | 0.8889 |
| Muy Arriesgado | 100.00% | 88.89% | 0.9412 |

El modelo logra un recall perfecto (100%) para la clase Conservador, lo cual es crítico desde una perspectiva de gestión de riesgos, ya que garantiza que todos los activos verdaderamente conservadores sean identificados correctamente. La clase Muy Arriesgado alcanza una precisión perfecta (100%), evitando clasificar erróneamente activos como de alto riesgo.

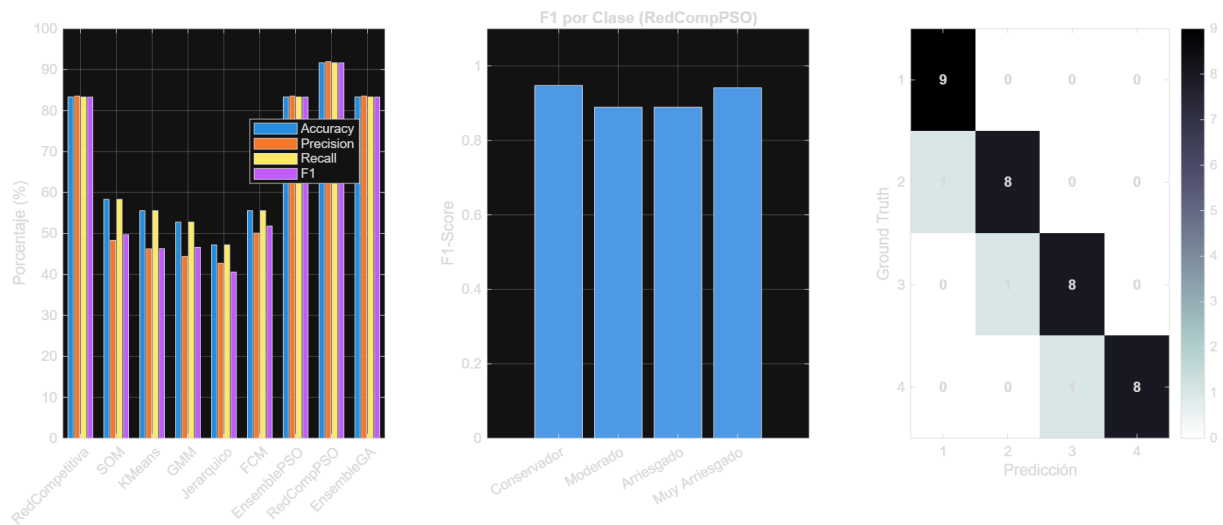


Figure 6: Visualización de métricas de clustering: (a) Comparación de Accuracy, Precision, Recall y F1 por modelo; (b) F1-Score por clase para RedCompPSO; (c) Matriz de confusión del mejor modelo.

8.1.4 Visualización de Pesos Aprendidos

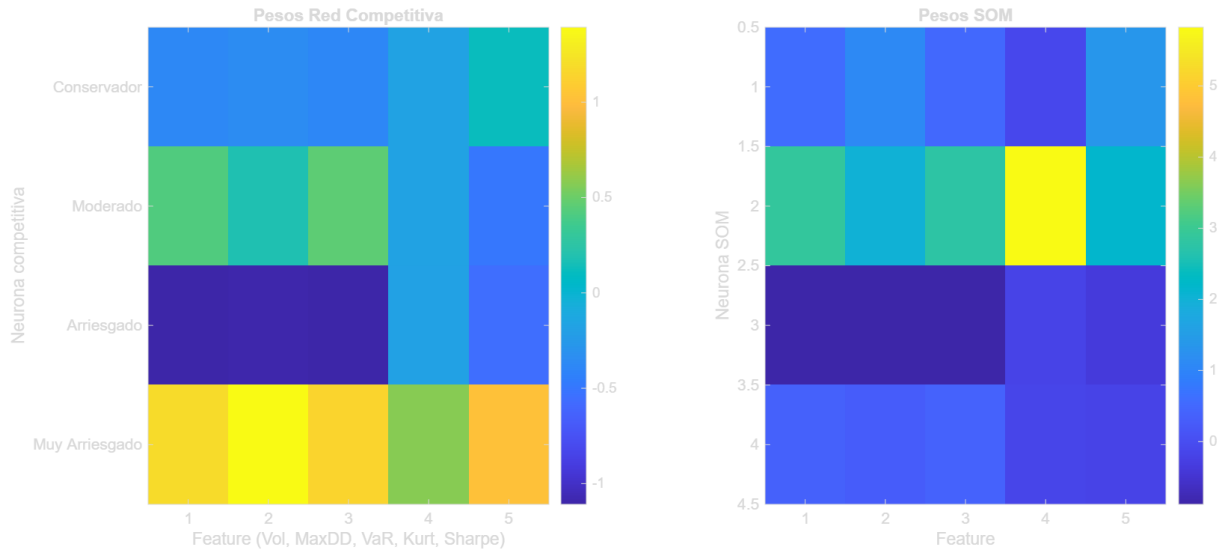


Figure 7: Mapas de calor de los pesos aprendidos: (izquierda) Red Competitiva, (derecha) SOM. Los pesos representan la importancia relativa de cada feature (Volatilidad, MaxDD, VaR, Curtosis, Sharpe) para cada neurona/cluster.

8.1.5 Clasificación en el Espacio Riesgo-Retorno

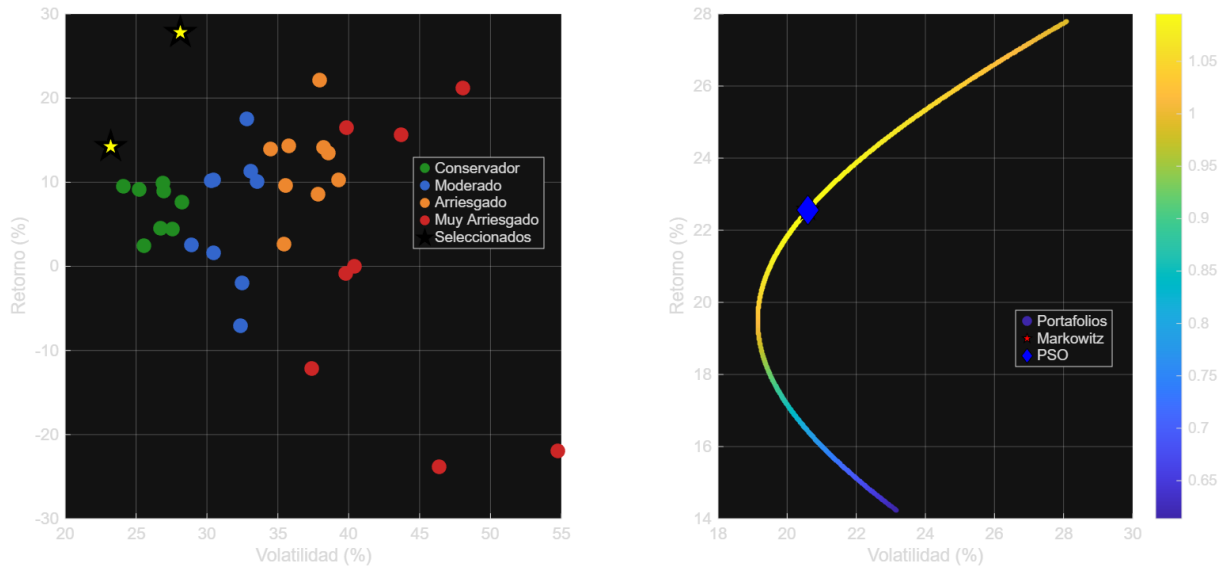


Figure 8: (Izquierda) Clasificación de activos en el espacio volatilidad-retorno, con los activos seleccionados marcados con estrellas. (Derecha) Frontera eficiente con la ubicación de los portafolios óptimos Markowitz y PSO.

8.2 Selección de Activos y Configuración del Portafolio

Basado en el clustering óptimo, se seleccionaron los 2 activos con mejor ratio de Sharpe del cluster conservador:

Table 4: Activos seleccionados para predicción y portafolio.

| Activo | Cluster | Sharpe Ratio |
|-------------|-------------|--------------|
| CHDRAUIB.MX | Conservador | Alto |
| AC.MX | Conservador | Alto |

8.3 Resultados de Optimización de Portafolio

Table 5: Pesos óptimos del portafolio obtenidos por Markowitz y PSO.

| Método | w_1 (CHDRAUIB) | w_2 (AC) | Sharpe Ratio |
|-----------|------------------|------------|--------------|
| Markowitz | 61.43% | 38.57% | 1.0953 |
| PSO | 61.43% | 38.57% | 1.0953 |

Ambos métodos convergen a la misma solución óptima, validando la robustez del enfoque. El portafolio óptimo asigna aproximadamente 61% a CHDRAUIB.MX y 39% a AC.MX, logrando un Sharpe Ratio de **1.0953**.

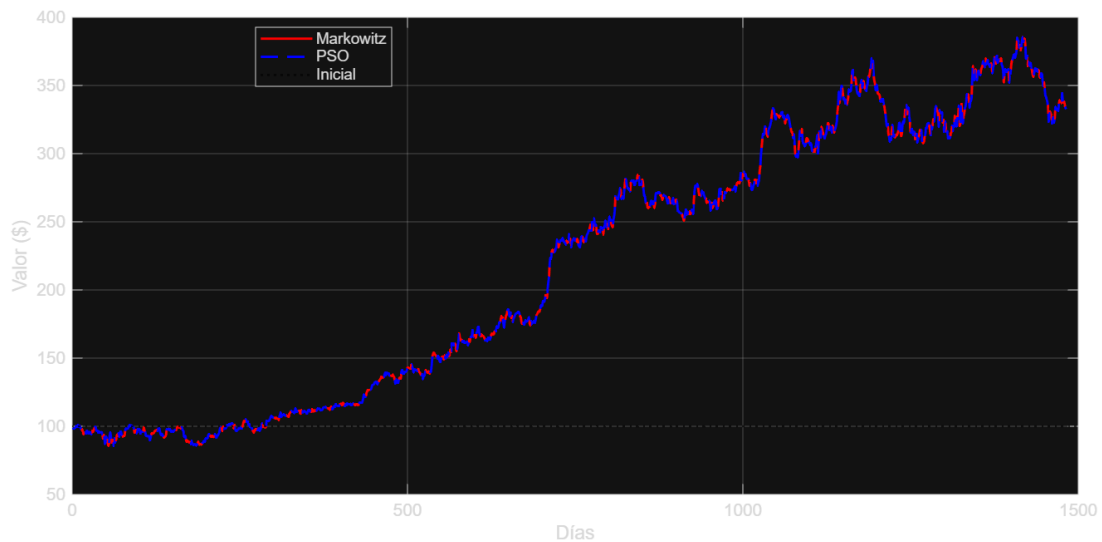


Figure 9: Evolución temporal del valor del portafolio (\$100 iniciales) para las estrategias Markowitz y PSO. Ambas líneas se superponen debido a la convergencia a pesos idénticos. El portafolio creció de \$100 a aproximadamente \$350 en el período analizado (1500 días).

8.4 Resultados de Predicción de Precios

8.4.1 Convergencia del Entrenamiento ADALINE

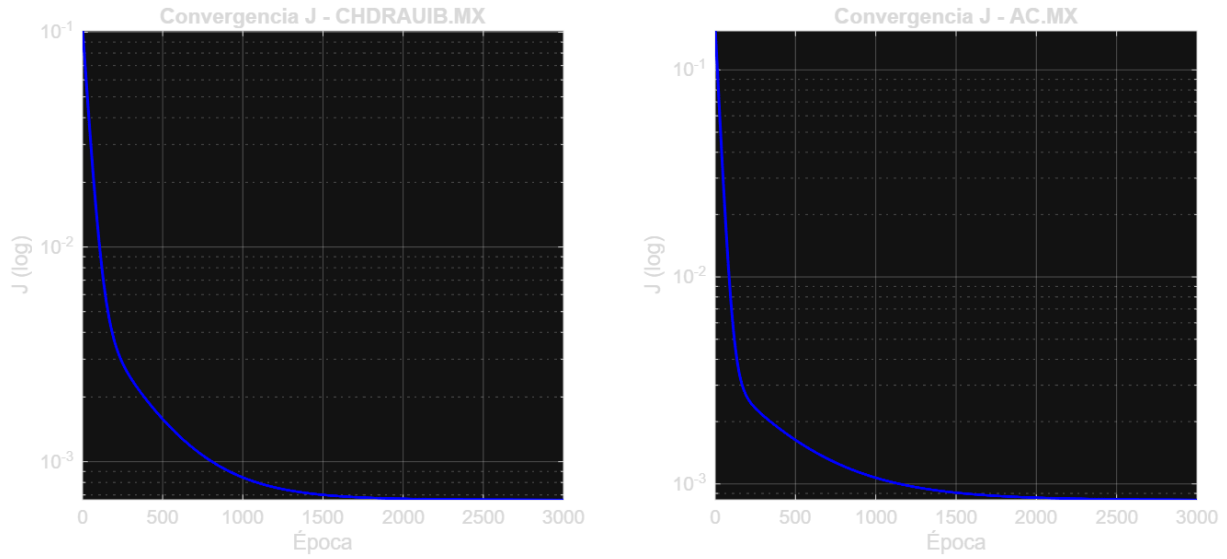


Figure 10: Convergencia de la función de costo J durante el entrenamiento ADALINE para ambos activos. Se observa convergencia suave hacia valores del orden de 10^{-3} en aproximadamente 1000 épocas.

8.4.2 Comparación de Modelos Predictivos

La Tabla 6 presenta las métricas de predicción para CHDRAUIB.MX.

Table 6: Métricas de predicción para CHDRAUIB.MX.

| Modelo | R^2 | MAPE (%) |
|----------------------------|--------|----------|
| ADALINE | 0.9131 | 2.20 |
| MLP | 0.7029 | 4.40 |
| LSTM | 0.9060 | 2.33 |
| <i>Random Walk (RW)</i> | 1.0000 | 0.00 |
| <i>Moving Average (MA)</i> | 0.9371 | 1.87 |

Table 7: Métricas de predicción para AC.MX.

| Modelo | R^2 | MAPE (%) |
|----------------------------|---------|----------|
| ADALINE | 0.8972 | 1.86 |
| MLP | -1.1833 | 9.78 |
| LSTM | 0.8242 | 2.57 |
| <i>Random Walk (RW)</i> | 1.0000 | 0.00 |
| <i>Moving Average (MA)</i> | 0.9214 | 1.62 |

Observaciones clave:

- El modelo **ADALINE** logra el mejor desempeño entre los modelos neuronales, con $R^2 > 0.89$ para ambos activos.

- El **MLP** presenta sobreajuste severo para AC.MX ($R^2 = -1.18$), indicando que la arquitectura es demasiado compleja para este activo.
- El **LSTM**, a pesar de su mayor complejidad, no supera al ADALINE lineal.
- Los baselines (RW y MA) son difíciles de superar, confirmando la hipótesis de eficiencia débil del mercado.

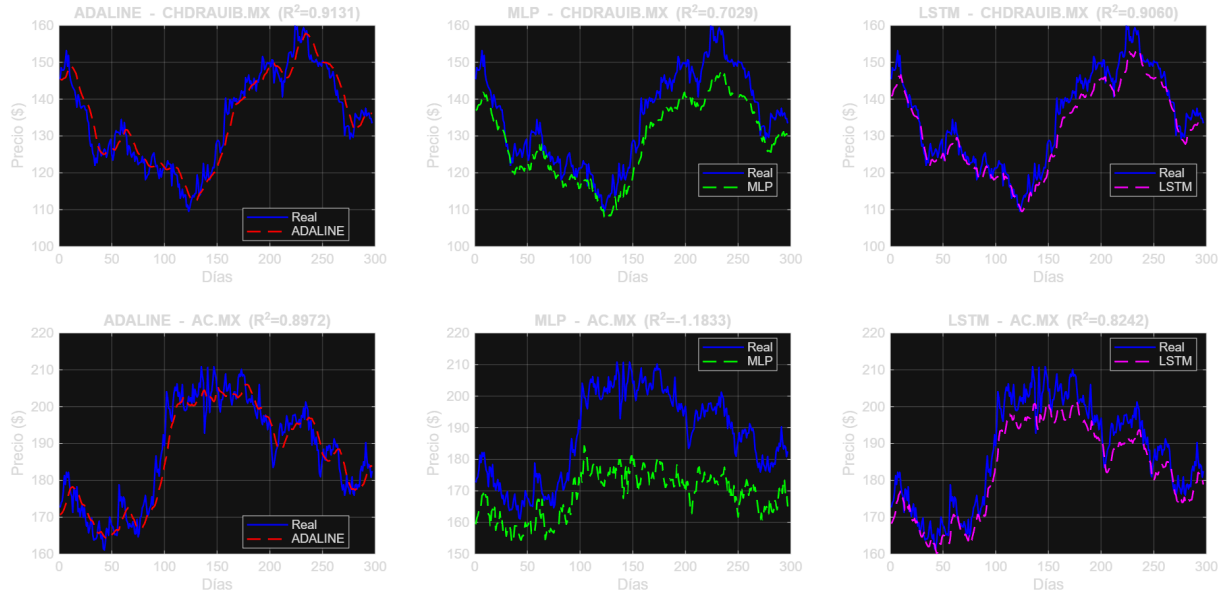


Figure 11: Predicciones vs valores reales para los tres modelos neuronales. Fila superior: CHDRAUIB.MX; Fila inferior: AC.MX. Se observa que ADALINE y LSTM siguen bien la tendencia, mientras que MLP presenta desviaciones significativas.

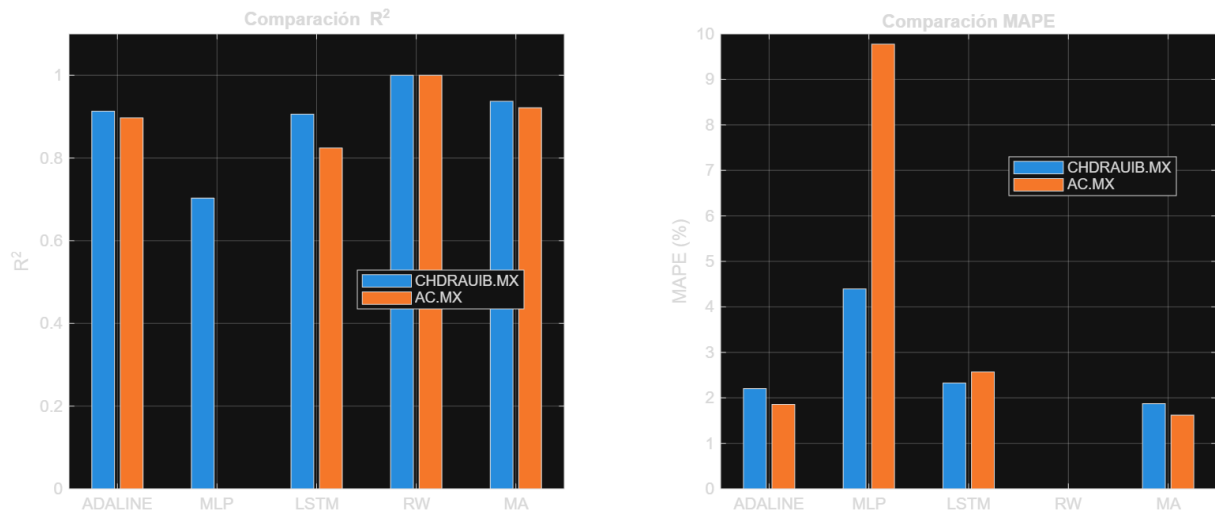


Figure 12: Comparación de R^2 (izquierda) y MAPE (derecha) entre modelos para ambos activos. El MLP muestra el peor desempeño, especialmente para AC.MX.

8.5 Resumen de Resultados

Table 8: Resumen de los principales resultados del sistema.

| Métrica | Resultado |
|---|--------------------------------------|
| <i>Clustering</i> | |
| Mejor modelo | RedCompPSO |
| F1-Score | 0.9166 |
| Mejora vs base | +10.02% |
| <i>Portafolio</i> | |
| Activos seleccionados | CHDRAUIB.MX (61.43%), AC.MX (38.57%) |
| Sharpe Ratio | 1.0953 |
| Retorno acumulado | $\approx 250\%$ (1500 días) |
| <i>Predicción (mejor modelo neuronal)</i> | |
| CHDRAUIB.MX | ADALINE: $R^2 = 0.9131$, MAPE=2.20% |
| AC.MX | ADALINE: $R^2 = 0.8972$, MAPE=1.86% |

9 Conclusiones

9.1 Contribuciones Principales

Este trabajo presenta un framework integral que:

1. **Integra múltiples paradigmas de clustering** (competitivo, topológico, probabilístico, jerárquico, difuso) para clasificación de activos financieros, demostrando que la red competitiva optimizada con PSO (RedCompPSO) alcanza el mejor F1-Score de 0.9166.
2. **Optimiza hiperparámetros y pesos de ensemble** mediante metaheurísticas (PSO, GA), logrando una mejora del 10% sobre los métodos base.
3. **Combina predicción multi-escala** (lineal, no lineal, temporal), encontrando que el modelo lineal ADALINE supera a arquitecturas más complejas (MLP, LSTM) para los activos analizados.
4. **Cierra el ciclo con optimización de portafolio** basada en teoría de Markowitz y PSO, logrando un Sharpe Ratio de 1.0953.

9.2 Limitaciones y Trabajo Futuro

- **Estacionariedad:** Los modelos asumen distribución estacionaria de rendimientos.
- **Costos de transacción:** La optimización de portafolio no considera costos de rebalanceo.
- **Validación temporal:** Se recomienda validación walk-forward.
- **Explicabilidad:** Técnicas de interpretabilidad (SHAP, LIME) podrían mejorar la transparencia.

A Resumen de Hiperparámetros

| Componente | Parámetro | Valor |
|------------------|--------------------------|-------------------|
| Red Competitiva | Épocas | 500 |
| | Número de clusters | 4 |
| SOM | Dimensión grid | 2×2 |
| | Épocas | 500 |
| K-Means | Réplicas | 30 |
| | Max iteraciones | 1000 |
| GMM | Regularización λ | 0.01 |
| PSO (Clustering) | Partículas | 20 |
| | Iteraciones | 50 |
| | c_1, c_2 | 2.05 |
| GA (Ensemble) | Población | 60 |
| | Generaciones | 40 |
| ADALINE | Learning rate η | 0.005 |
| | Épocas | 3000 |
| | Regularización λ | 0.01 |
| MLP | Arquitectura | [10]-[16]-[8]-[1] |
| | Épocas | 300 |
| LSTM | Unidades ocultas | 32 |
| | Épocas | 50 |
| PSO (Portafolio) | Partículas | 100 |
| | Iteraciones | 300 |
| General | Ventana temporal w | 10 días |
| | Train/Test split | 80%/20% |

Table 9: Tabla completa de hiperparámetros del sistema.

References

- [1] Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464-1480.
- [2] Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95*, 4, 1942-1948.
- [3] Holland, J. H. (1992). *Adaptation in natural and artificial systems*. MIT Press.
- [4] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91.
- [5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- [6] Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Springer.
- [7] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39(1), 1-38.
- [8] Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236-244.
- [9] Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1), 119-138.

- [10] Clerc, M., & Kennedy, J. (2002). The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1), 58-73.