

Machine Learning aplicado al Trading

Modelo XgBoost de clasificación binaria aplicado a trading de corto plazo en AAPL (2019-2025)

Ignacio Vergara



Objetivos del Proyecto y Target del Modelo

- Explorar como afecta el uso de diferentes indicadores técnicos en el modelo para determinar cuales son más cruciales en las estrategias de trading
- Crear un modelo capaz de anticipar subidas del mercado y quedarnos fuera en las caídas para batir a la estrategia referencia buy and hold(estar siempre comprado y no vender)
- Probar varios modelos para aprender sus diferencias y concluir cuál es el mejor para nuestro caso de estudio

Target

Para identificar días con subidas significativas y eliminar días laterales de mucho ruido, defino como target un umbral de retorno objetivo que determinan la clase positiva:(por ejemplo 0.5% diario).

Features



Momentum(ROC)

Retornos diarios y Rate of Change (ROC) de 10 días para capturar la tendencia direccional del precio.



Volatilidad

Desviación estándar de 10 días y rango diario (high-low) para cuantificar el riesgo y la incertidumbre del mercado.



Volumen Relativo

Volumen relativo como indicador de participación del mercado y confirmación de movimientos de precio.



Media móvil de filtro

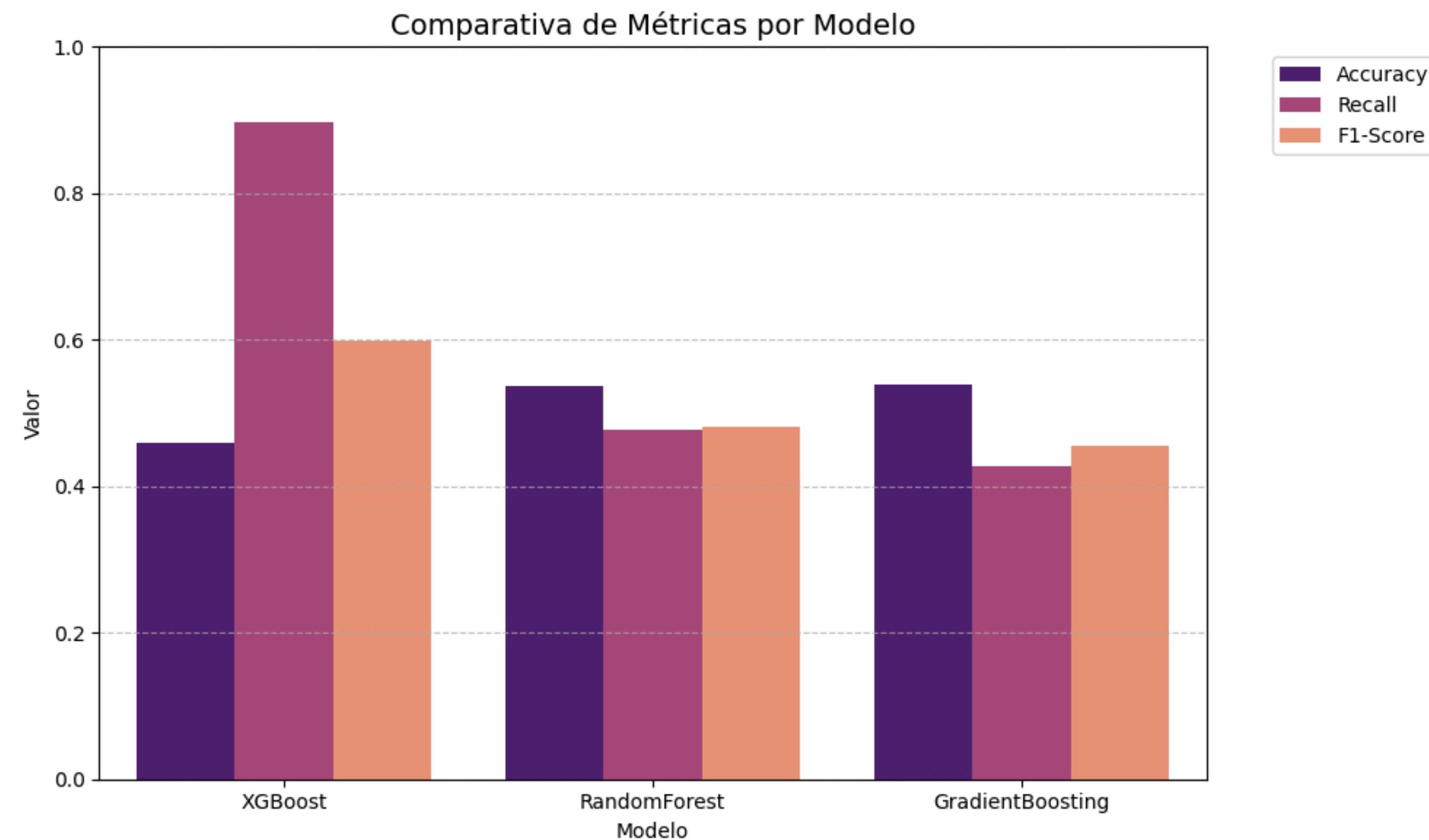
Media móvil exponencial de 30 días (EMA30) como confirmación adicional de tendencia subyacente.

Money Flow Index

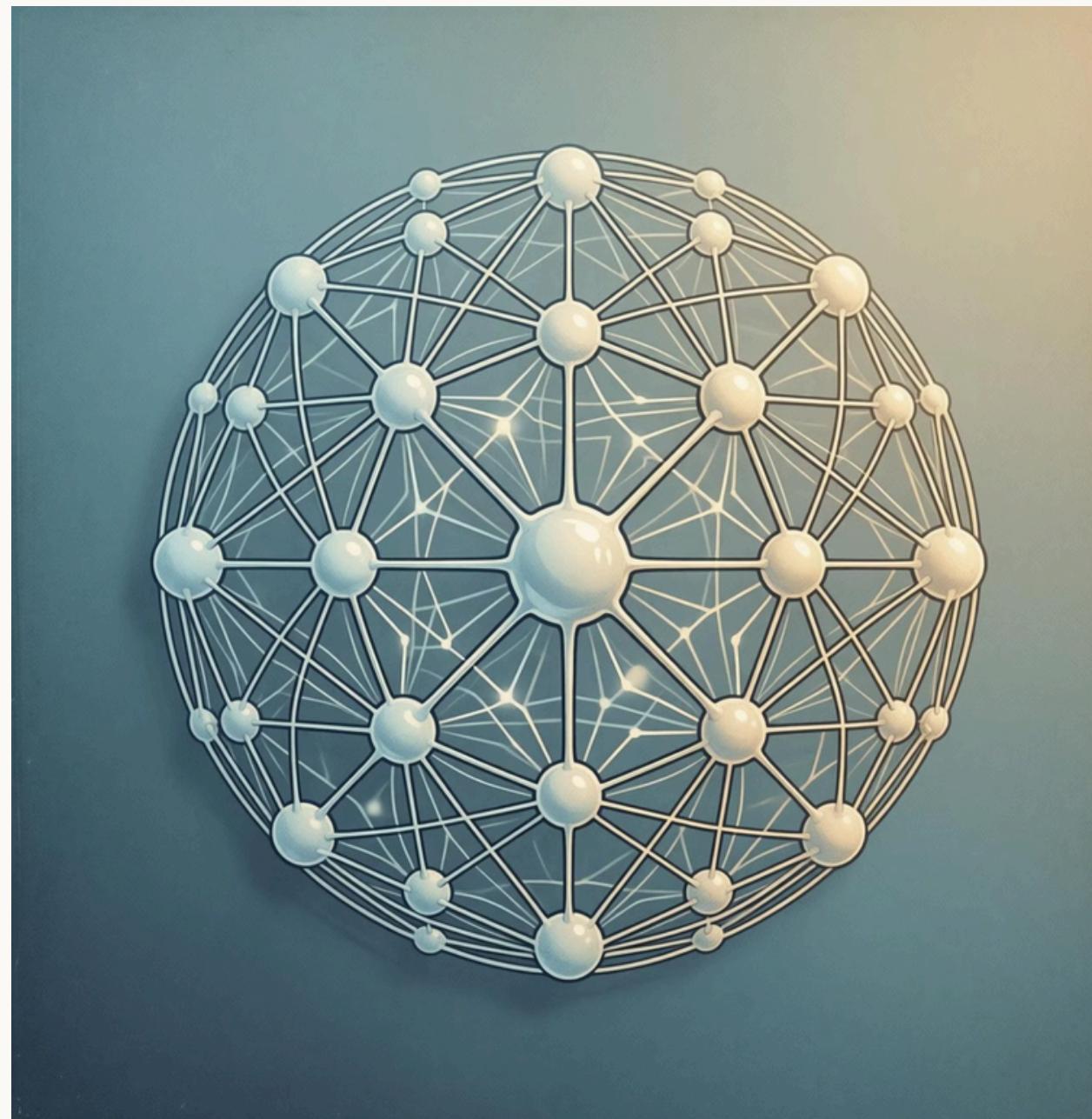
Combina precio y volumen para medir la presión de compra

Modelos

Modelos probados con Optuna: Logistic Regression, RandomForest y GradientBoostin. Gano XgBoost



Arquitectura del Modelo



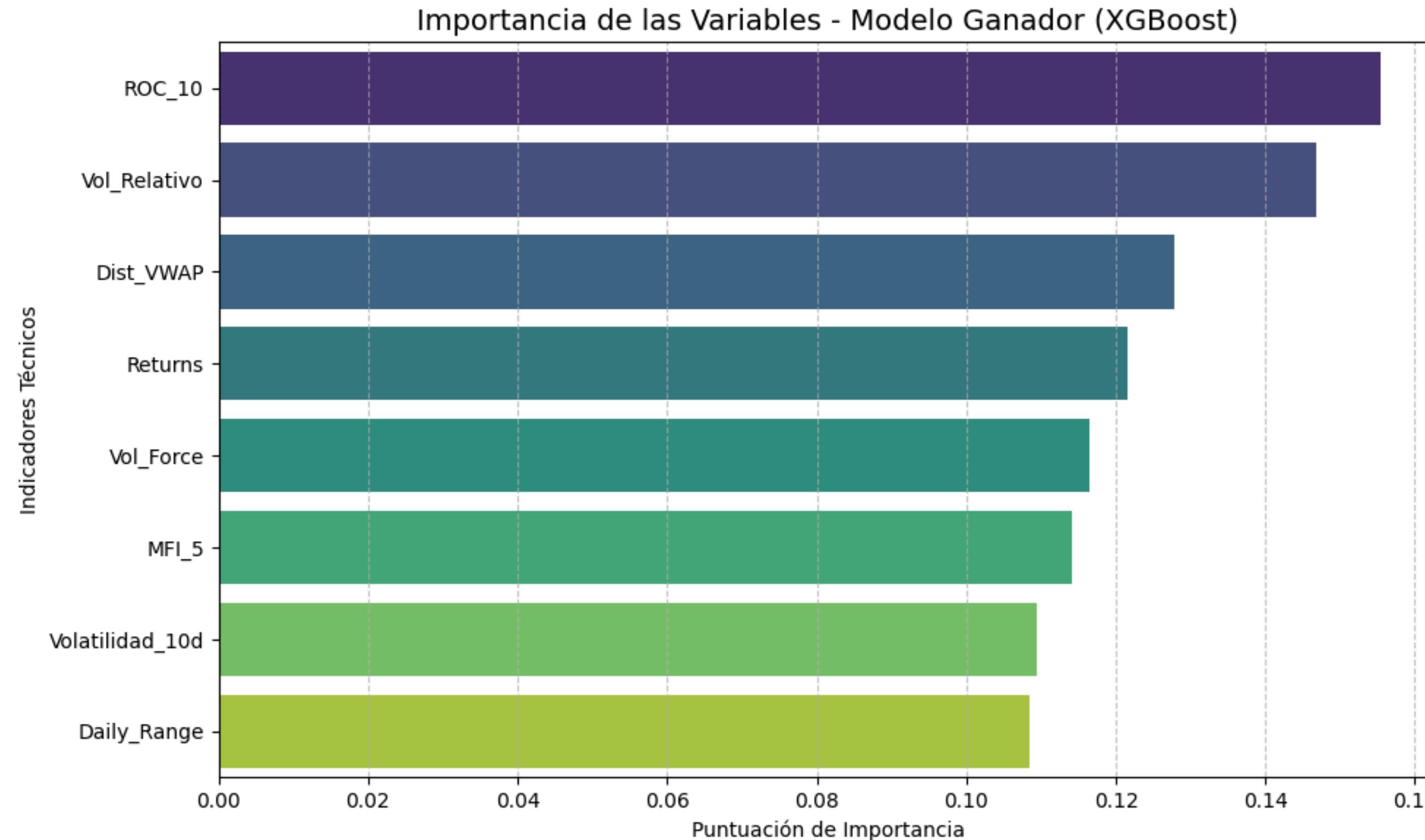
XGBoost Classifier

Elegido por su robustez ante desbalance de clases y capacidad de capturar relaciones no lineales. Además la función de feature importance me ayudó seleccionar los indicadores finales.

Hiperparámetros Clave

- **n_estimators:** 250 árboles para balance entre precisión y overfitting
- **max_depth:** 6 niveles para controlar complejidad del modelo
- **learning_rate:** 0.05 para convergencia gradual y estable

Feature importance



Métricas de Machine Learning

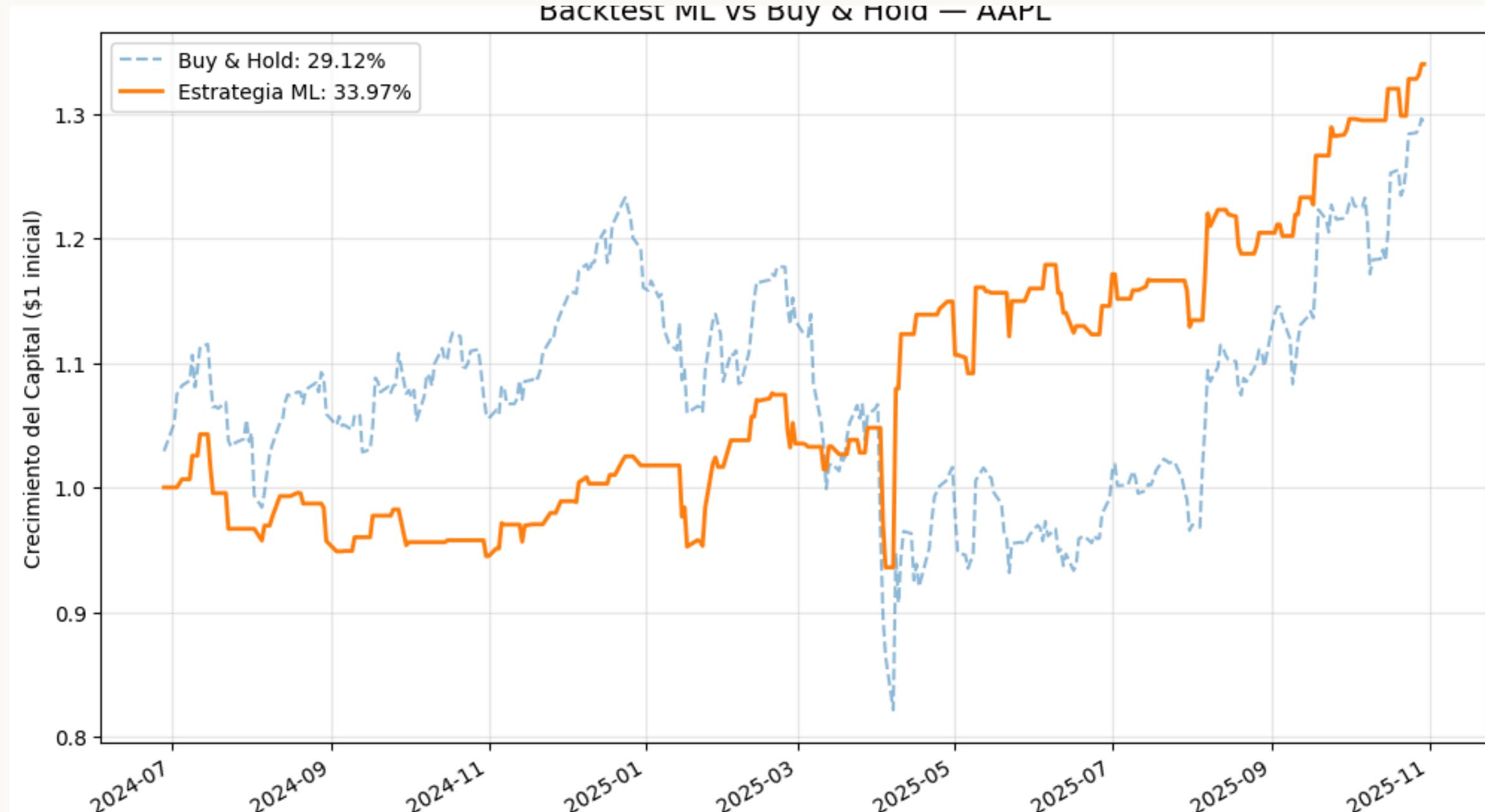
Métrica	Valor
Precisión (Clase 1)	0.48
Recall (Clase 1)	0.41
F1-score (Clase 1)	0.44
F1-score (Clase 0)	0.59

Interpretación Crítica

Las métricas tradicionales de ML muestran valores moderados, especialmente para la clase positiva (subidas significativas).

Sin embargo: Estas métricas no reflejan completamente el potencial de la estrategia financiera. La calidad del trade importa más que la tasa de aciertos.

Desempeño Financiero en Backtest



33.97%

Retorno Total

Estrategia ML vs 29.12% Buy & Hold

1.77

Sharpe Ratio
Ratio que relaciona
rendimiento y riesgo
implicito

-13%

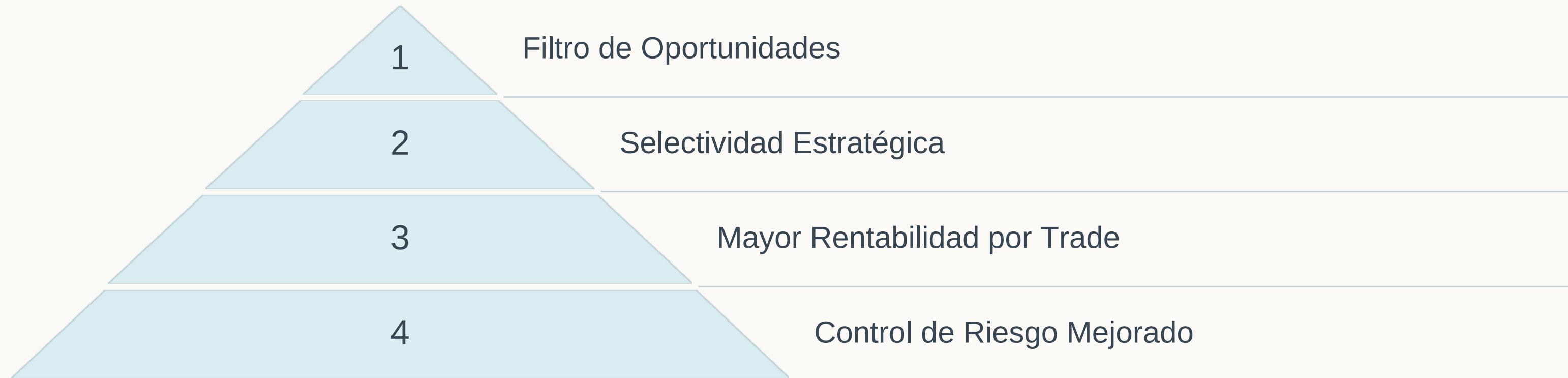
Max Drawdown
vs -32% Buy & Hold, control
de riesgo mejorado

130

Trades Totales

Selectividad alta, calidad
sobre cantidad

Interpretación: Calidad sobre Cantidad



Función del Modelo

El modelo no busca predecir todas las subidas, sino identificar las oportunidades con mejor relación riesgo-retorno y quedarse fuera en los cambios de tendencia. Actúa como un sistema de filtrado inteligente.

Umbral y Rentabilidad

Umbrales más exigentes generan menos señales pero trades más rentables. La baja frecuencia se compensa con mayor calidad y consistencia.

- ❑ Métricas ML bajas no implican estrategia inútil: el valor está en maximizar el retorno esperado por operación, no en maximizar métricas como accuracy.

Conclusiones y Aplicabilidad

ML como Filtro, No es un Oráculo

Machine learning es útil para filtrar operaciones de alta calidad, no para predecir todos los movimientos del mercado ni basarnos en el exclusivamente. El enfoque es selectivo y estratégico.

Target Return

Es un parámetro muy determinante. El umbral de retorno objetivo determina la dificultad del problema de clasificación y la rentabilidad esperada.

Calidad sobre cantidad

Un enfoque práctico de trading que prioriza el retorno esperado por operación sobre la tasa de aciertos bruta.

ML + Análisis Técnico

El modelo puede aprender patrones y utilizarlos para generar sistemas rentables.
Habrá que hacer backtest real en producción

