

InelidaMarketScan — Análisis Científico de los Mercados Financieros con Machine Learning

Investigación en Finanzas Cuantitativas: ICT, Ichimoku, XGBoost y Cálculo Estocástico Por **Didier Vally** / Reuniware Systems Versión 1.0.6 — 2026-07-19
□ PyPI: <https://pypi.org/project/inelida-marketscan> | GitHub: <https://github.com/reuniware/InelidaMarketScanner>

Resumen

Este documento presenta los resultados de 6 meses de investigación en finanzas cuantitativas aplicadas al trading algorítmico. El framework InelidaMarketScan combina el análisis técnico ICT (Inner Circle Trader) e Ichimoku multi-timeframe con un pipeline de Machine Learning XGBoost (123 features) y funciones de cálculo estocástico (Ornstein-Uhlenbeck, Hurst, GARCH, Monte Carlo, Kelly). El modelo v18 alcanza un 64,0% de precisión en validación cruzada sobre 1.210 operaciones históricas. El análisis de 27 modelos sucesivos revela que las features de Ichimoku (distancias Kijun/Tenkan, cruces T/K) son más predictivas que los patrones ICT (FVG, Order Blocks). La simulación de capital con filtrado $ML\% \geq 50\%$ muestra un factor de beneficio de 2,27 en datos in-sample, mientras que la validación out-of-sample identifica un techo en 1,173. Este documento constituye una referencia científica completa sobre la intersección entre ICT, Ichimoku y Machine Learning.

1. Introducción

1.1 Contexto y Problemática

El trading algorítmico se basa en la identificación de patrones recurrentes en las series temporales financieras. Los conceptos ICT (Inner Circle Trader) desarrollados por Michael Huddleston describen el comportamiento institucional de los mercados: barridos de liquidez, Fair Value Gaps, Order Blocks, Market Structure Shifts. Paralelamente, el indicador Ichimoku Kinko Hyo proporciona una lectura multi-timeframe del equilibrio del mercado. Este proyecto explora la fusión de estos dos paradigmas mediante un escáner automatizado acoplado a un modelo XGBoost.

1.2 Objetivos de Investigación

- Desarrollar un escáner automático que combine ICT e Ichimoku en 8 timeframes (M5 a MN).
- Construir un pipeline ML con 123 features nombradas extraídas del escáner.
- Validar el poder predictivo mediante backtest histórico (2025-2026) y prueba out-of-sample.
- Integrar funciones de finanzas cuantitativas (OU, Hurst, GARCH, Monte Carlo, Kelly).
- Alcanzar un factor de beneficio $> 1,5$ en condiciones reales.

2. Fundamentos Teóricos ICT

Los conceptos ICT modelan el comportamiento de las instituciones financieras (bancos, hedge funds) que manipulan los niveles de precios para acumular posiciones. Los 8 conceptos siguientes son detectados automáticamente por el escáner:

Concepto ICT	Descripción y Detección
Barridos de Liquidez	Ruptura breve de un nivel clave (Asian High/Low, London High/Low) para activar stops antes de la inversión. Detectado en H1 con confirmación de retroceso.
Fair Value Gaps (FVG)	Brecha entre 3 velas consecutivas donde el precio no negoció. Detectado en 6 TFs (M5 a D1) con porcentaje de mitigación.
Order Blocks (OB)	Última vela antes de un impulso direccional — zona donde las instituciones colocaron órdenes. Detectado vía ATR(14) en 6 TFs.
Breaker Blocks (BRK)	Order Block roto que se invierte en soporte/resistencia opuesto. Derivado de los OB existentes.
Market Structure Shift (MSS/BOS)	Cambio en la estructura del mercado: Break of Structure y Change of Character en 6 TFs.
Volumen HVN/LVN	High Volume Node y Low Volume Node basados en el tick volume de MT5. Detectado en 6 TFs.
Reversal Pipeline	Cadena completa barrido → retroceso → confirmación con contador de velas post-barrido.
Extensiones de Fibonacci	Extensiones +1,618 a +4,0 basadas en el rango asiático para objetivos de TP.

3. Arquitectura Ichimoku Multi-TF

El escáner Diamond analiza 8 timeframes simultáneamente. Para cada TF, se calculan 5 componentes Ichimoku:

Timeframe	Tenkan	Kijun	Kumo	T/K Cross	Flat History
M5	☐	☐	—	☐	—
M15	☐	☐	—	☐	—
M30	☐	☐	—	☐	—
H1	☐	☐	☐	☐	☐
H4	☐	☐	☐	☐	☐
D1	☐	☐	☐	☐	☐
W1	☐	☐	☐	☐	☐
MN	☐	☐	☐	☐	☐

4. Arquitectura del Escáner Diamond

El Diamond Scanner agrega 109 criterios de puntuación en 10 categorías:

Categoría	Criterios	Ejemplos
Ichimoku HI-D1	24	Distancias Kijun/Tenkan, cruces T/K, posiciones Kumo, Flat bars
Ichimoku W1-MN	8	Kijun W1/MN, Kumo W1/MN, color de nube, Doble memoria
FVG (6 TFs)	12	FVG bear/bull M5 a D1, % de mitigación, prioridad temporal
Order Blocks (6 TFs)	12	OB bear/bull M5 a D1, posición ARRIBA/DEBAJO/EN
MSS/BOS (6 TFs)	18	Régimen, BOS, CHoCH en M5 a D1
Volumen HVN/LVN (6 TFs)	12	Pico de volumen, HVN, LVN en M5 a D1
Breaker Blocks (6 TFs)	12	BRK bear/bull M5 a D1
Barridos Asian/London	4	Barrido AH/AL, barrido London H/L
Reversal Pipeline	5	Barrido→retroceso→confirmación, FVG post-barrido, retest
Kijun M30/M15/M5	2	Cruce direccional Kijun M30/M15/M5

Puntuación máxima: 109 criterios (107 sin MN)

5. Pipeline de Machine Learning

El pipeline ML transforma los resultados del escáner en probabilidad de ganancia en 3 etapas:

Etap	Herramienta	Detalles
1	ml_labeler.py	Extracción de features del escaneo Diamond (BD / CSV manual / backtest). Genera CSV etiquetado con columna outcome (1=ganada, 0=perdida).
2	ml_trainer.py	Entrenamiento XGBoost con validación cruzada 5-fold, scale_pos_weight para clases desequilibradas, búsqueda grid de hiperparámetros.
3	ml_predictor.py	Inferencia en tiempo real: extract_features() → predict_proba() → ML% mostrado en el escaneo en vivo. MLPredictor simplificado: 1 joblib.load en vez de 5.

6. Modelo XGBoost v18 — Configuración

El modelo v18 es el modelo de producción activo. Entrenado en 1.210 operaciones históricas con 123 features nombradas (incluyendo GARCH y estocástico).

Hiperparámetro	Valor
max_depth	6
learning_rate	0.03
n_estimators	200
subsample	0.8
min_child_weight	3
gamma	0.1
reg_alpha	0.5 (L1)
reg_lambda	1.0 (L2)
scale_pos_weight	auto (basado en ratio win/loss)

123 features nombradas incluyen: scores/sesgo (6), Kijun 6 TFs (18), Tenkan 6 TFs (18), cruces T/K (6), Kumo (6), cruces M5/M15/M30 (12), Flat bars (2), FVGs activos (1), Order Blocks (6), Barridos (5), Volumen (6), MSS/BOS (6), Breaker Blocks (6), Compresión (1), Estocástico (8: OU, Hurst, Monte Carlo, GARCH), Temporal (3), Tipo de activo (3).

7. Historial de Modelos

27 modelos XGBoost han sido entrenados y evaluados desde el inicio del proyecto. Aquí el historial de las versiones principales:

Modelo	Datos	Features	Operaciones	CV	OOS PF
v1	6 principales	36	852	61.0%	—
v4	13 símbolos	111	1.210	62.9%	1.26*
v7	13 símbolos	25 mejores	2.923	60.1%	1.173
v13	4 modelos/activo	25	2.923	53-64%	4.000
v15	2025-2026	25 + OU/Hurst	5.009	59.8%	—
v18	13 símbolos	123 (GARCH)	1.210	64.0%	0.503 F1

8. Importancia de Features (v4)

El análisis de importancia de features revela qué criterios pesan más en las decisiones del modelo:

Rank	Feature	Ganancia	Insight
1	is_dxy	14.2	El dólar sigue una lógica radicalmente diferente a los pares forex
2	rr	10.0	El ratio risk/reward teórico predice el resultado real
3	tkx_h1	9.2	El cruce Tenkan/Kijun H1 es el mejor predictor individual
4	kj_h1	6.9	La distancia a la Kijun H1 captura el equilibrio a corto plazo
5	s1	6.4	El nivel de stop-loss es tan importante como el take-profit
6	day_wednesday	6.4	Los miércoles (FOMC) muestran comportamiento estadísticamente distinto
7	d_kj_h4	6.3	La distancia a la Kijun H4 indica la tendencia subyacente
8	d_kj_d1	6.0	La distancia a la Kijun D1 ancla el contexto macro
9	tkx_h4	6.0	El cruce T/K H4 confirma la tendencia de sesión
10	is_forex	5.9	Los pares FX tienen comportamiento propio vs índices/metales

9. Validación Out-of-Sample

La validación OOS (entrenamiento Enero-Junio 2026, prueba Julio 2026) es la prueba decisiva. El modelo v4 muestra un PF in-sample de 2,27 en el umbral $ML\% \geq 50\%$, pero esto incluye data leakage (el modelo vio el 80% de los datos de prueba durante el entrenamiento). El verdadero PF OOS del v7 es 1,173 — un edge estadísticamente significativo pero modesto. El modelo v13 (por activo) alcanza $PF=4,000$ en 15 operaciones, pero la muestra es demasiado pequeña para ser fiable. Conclusión: el ML proporciona un edge real ($PF > 1,0$), pero el techo actual es $\sim 1,17$ con arquitectura unificada.

10. Simulación de Capital — Filtrado $ML\%$

Simulación sobre 1.210 operaciones históricas con capital inicial de 10.000 € y riesgo del 1% por operación:

Umbral ML%	Operaciones	Tasa de Acierto	PF (RR)	G&P (10k€)
Sin filtro	1.210	37.6%	1.02	+1.442 €
ML% $\geq$ 45%	661	54.3%	1.91 $\square$	+27.542 €
ML% $\geq$ 50%	504	59.7%	2.27 $\square\square$	+25.842 €
ML% $\geq$ 55%	327	67.3%	2.80	+19.242 €
ML% $\geq$ 60%	170	78.8%	3.59	+9.340 €
ML% $\geq$ 70%	71	88.7%	1.64	+514 €

11. Finanzas Cuantitativas y Cálculo Estocástico

Se han integrado 6 funciones de finanzas cuantitativas en el escáner y el pipeline ML. Estas funciones calculan métricas de riesgo, volatilidad y régimen de mercado:

Función	Objetivo	Feature ML	Rank v18
Ornstein-Uhlenbeck	Detecta la sobreextensión del precio respecto a su media a largo plazo. Score 0-100%: >80% = retroceso probable.	ou_score	No probado
Exponente de Hurst	Clasifica el régimen de mercado: $H < 0.45$ = RANGO, $H > 0.55$ = TENDENCIA.	hurst_h	No probado
GARCH(1,1)	Modela la volatilidad condicional. 4 features: persistencia, vol_largo_plazo, vida_media, vol_pronóstico.	garch_persistence	#14
Monte Carlo	Simula 10.000 trayectorias de precio. Calcula P(SL alcanzado) y P(TP alcanzado).	mc_p_sl, mc_p_tp	No probado
Criterio de Kelly	Dimensionamiento óptimo de posiciones por tipo de activo. Ganancia=0.00 (redundante con SL/TP ya optimizados).	kelly_full	#124
Ruina FTMO	Probabilidad de ruina de una cuenta de prop firm con drawdown del 10% en 100 operaciones simuladas.	ftmo_ruin_prob	No probado

12. Criterio de Kelly por Tipo de Activo

El Criterio de Kelly utiliza tasas de acierto reales por tipo de activo (del backtest) en lugar de un `win_rate=0.40` uniforme. Resultados:

Tipo de Activo	Tasa de Acierto	Kelly Full	Kelly Half
DXY	73%	46.0%	23.0%
XAUUSD	42%	13.0%	6.5%
Forex	31%	3.5%	1.8%
Índices	38%	7.0%	3.5%
Cripto	35%	2.5%	1.3%

13. MLPredictor Simplificado (19/07/2026)

El MLPredictor cargaba anteriormente el mismo modelo 5 veces ($1\times$ por método de predicción). La simplificación del 19/07/2026 lo redujo a un solo `joblib.load()`. Resultado: 30 líneas de código eliminadas, tiempo de carga dividido por 4, y el `_ASSET_MODEL_MAP` redundante eliminado. El modelo v18 es ahora el modelo unificado único para todos los tipos de activos.

14. Corrección de Errores — 19/07/2026

Una revisión de código exhaustiva el 19/07/2026 identificó y corrigió 4 errores:

#	Gravedad	Archivo	Descripción
1	CRÍTICO	diamond_scanner.py	Código muerto después de return en ml_available — los símbolos nunca se activaban. initialize() no establecía _initialized = True.
2	ALTO	stochastic_analytics.py	GARCH: si no se encontraba un par (α, β) válido, fallback silencioso con parámetros inventados. Añadida verificación best_ll <= -1e100.
3	MEDIO	stochastic_analytics.py	Monte Carlo: el fallback de volatilidad (log-returns) se calculaba pero nunca se usaba cuando el sigma OU era cero.
4	BAJO	stochastic_analytics.py	FTMO: pnl_avg/roi_avg fuera del bloque try — riesgo de NameError si se añadía código entre el cálculo y el return.

15. Matriz de Decisión Final

La matriz de decisión combina la puntuación Diamond, el ML%, los guardarraíles y el contexto macro:

Condición	Acción	Justificación
ML% $\geq$ 60% + TKx HI alineado + sin HIGH NEWS	☐ GO	Triple confirmación: ML, Ichimoku, macro
ML% $\geq$ 60% + TKx HI conflicto	☐ ESPERAR	El ML es contradicho por el cruce T/K HI
ML% 50-60% + TKx HI alineado	☐ PRECAUCIÓN	Edge ML moderado, necesita confirmación adicional
ML% < 50%	☐ NO GO	El modelo no tiene edge estadístico
HIGH NEWS DAY + BULL	☐ ESPERAR	Los días macro (≥ 2 eventos mayores) rompen correlaciones
Miércoles (día FOMC) + Forex	☐ Riesgo Alto	Los miércoles muestran comportamiento estadísticamente distinto

16. Conclusión y Perspectivas

El proyecto InelidaMarketScan demuestra que un pipeline ML entrenado con features Ichimoku puede generar un edge estadístico significativo (PF OOS = 1,173). Los 4 errores críticos corregidos el 19/07/2026 refuerzan la fiabilidad del escáner en vivo. Los 27 modelos sucesivos revelaron ideas clave: las features Ichimoku son más predictivas que los patrones ICT, el tipo de activo (is_dxy, is_forex) es un predictor principal, y las funciones de cálculo estocástico (GARCH, OU, Hurst) proporcionan señal complementaria de volatilidad. El próximo salto de rendimiento requerirá un dataset más grande (2024-2026, >5.000 operaciones/activo) y potencialmente arquitecturas de deep learning.

Referencias

- Huddleston, M. — Inner Circle Trader (ICT) Concepts, 2022-2024
- Chen, T. & Guestrin, C. — XGBoost: A Scalable Tree Boosting System, KDD 2016
- Bollerslev, T. — Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 1986
- Hurst, H.E. — Long-term Storage Capacity of Reservoirs, Trans. ASCE, 1951
- Uhlenbeck, G.E. & Ornstein, L.S. — On the Theory of Brownian Motion, Physical Review, 1930
- Kelly, J.L. — A New Interpretation of Information Rate, Bell System Technical Journal, 1956
- Hosoda, G. — Ichimoku Kinko Hyo, 1968