

InelidaMarketScan — Analyse Scientifique des Marchés Financiers avec Machine Learning

Recherche en Finance Quantitative : ICT, Ichimoku, XGBoost & Calcul Stochastique Par **Didier Vally** / Reuniware Systems Version 1.0.6 — 2026-07-19
□ PyPI: <https://pypi.org/project/inelida-marketscan> | GitHub: <https://github.com/reuniware/InelidaMarketScanner>

Résumé

Ce document présente les résultats de 6 mois de recherche en finance quantitative appliquée au trading algorithmique. Le framework InelidaMarketScan combine l'analyse technique ICT (Inner Circle Trader) et Ichimoku multi-timeframe avec un pipeline de Machine Learning XGBoost (123 features) et des fonctions de calcul stochastique (Ornstein-Uhlenbeck, Hurst, GARCH, Monte Carlo, Kelly). Le modèle v18 atteint 64.0% de précision en cross-validation sur 1 210 trades historiques. L'analyse de 27 modèles successifs révèle que les features Ichimoku (distances Kijun/Tenkan, croisements T/K) sont plus prédictives que les patterns ICT (FVG, Order Blocks). La simulation de capital avec filtrage $ML\% \geq 50\%$ montre un profit factor de 2.27 sur données in-sample, tandis que la validation out-of-sample identifie un plafond à 1.173. Ce document constitue une référence scientifique complète sur l'intersection entre ICT, Ichimoku et Machine Learning.

1. Introduction

1.1 Contexte et Problématique

Le trading algorithmique repose sur l'identification de patterns récurrents dans les séries temporelles financières. Les concepts ICT (Inner Circle Trader) développés par Michael Huddleston décrivent le comportement institutionnel des marchés : sweeps de liquidité, Fair Value Gaps, Order Blocks, Market Structure Shifts. Parallèlement, l'indicateur Ichimoku Kinko Hyo fournit une lecture multi-timeframe de l'équilibre du marché. Ce projet explore la fusion de ces deux paradigmes via un scanner automatisé couplé à un modèle XGBoost.

1.2 Objectifs de Recherche

- Développer un scanner automatique combinant ICT et Ichimoku sur 8 timeframes (M5 à MN).
- Construire un pipeline ML avec 123 features nommées extraites du scanner.
- Valider le pouvoir prédictif via backtest historique (2025-2026) et test out-of-sample.
- Intégrer des fonctions de finance quantitative (OU, Hurst, GARCH, Monte Carlo, Kelly).
- Atteindre un profit factor > 1.5 en conditions réelles.

2. Fondements Théoriques ICT

Les concepts ICT modélisent le comportement des institutions financières (banques, hedge funds) qui manipulent les niveaux de prix pour accumuler des positions. Les 8 concepts suivants sont détectés automatiquement par le scanner :

Concept ICT	Description et Détection
Sweeps de Liquidité	Cassure brève d'un niveau clé (Asian High/Low, London High/Low) pour déclencher les stops avant inversion. Détecté sur H1 avec confirmation de retournement.
Fair Value Gaps (FVG)	Écart entre 3 bougies consécutives où le prix n'a pas négocié. Détecté sur 6 TFs (M5 à D1) avec pourcentage de mitigation.
Order Blocks (OB)	Dernière bougie avant une impulsion directionnelle — zone où les institutions ont placé leurs ordres. Détecté via ATR(14) sur 6 TFs.
Breaker Blocks (BRK)	Order Block cassé qui s'inverse en support/résistance opposé. Dérivé des OB existants, aucun appel MT5 supplémentaire.
Market Structure Shift (MSS/BOS)	Changement de structure de marché : Break of Structure et Change of Character sur 6 TFs.
Volume HVN/LVN	High Volume Node et Low Volume Node basés sur le tick volume MT5. Détecté sur 6 TFs.
Reversal Pipeline	Chaîne complète sweep → retournement → confirmation avec compteur de barres post-sweep.
Fibonacci Extensions	Extensions +1.618 à +4.0 basées sur le range asiatique pour les objectifs de TP.

3. Architecture Ichimoku Multi-TF

Le scanner Diamond analyse 8 timeframes simultanément. Pour chaque TF, 5 composants Ichimoku sont calculés :

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

4. Architecture du Scanner Diamond

Le Diamond Scanner agrège 109 critères de scoring répartis en 10 catégories :

Catégorie	Critères	Exemples
Ichimoku HI-D1	24	Kijun/Tenkan distances, T/K crosses, Kumo positions, Flat bars
Ichimoku W1-MN	8	Kijun W1/MN, Kumo W1/MN, Cloud couleur, Double mémoire
FVG (6 TFs)	12	FVG bear/bull M5 à D1, mitigation %, priorité temporelle
Order Blocks (6 TFs)	12	OB bear/bull M5 à D1, position ABOVE/BELOW/AT
MSS/BOS (6 TFs)	18	Régime, BOS, CHoCH sur M5 à D1
Volume HVN/LVN (6 TFs)	12	Volume spike, HVN, LVN sur M5 à D1
Breaker Blocks (6 TFs)	12	BRK bear/bull M5 à D1
Asian/London Sweeps	4	AH/AL sweep, London H/L sweep
Reversal Pipeline	5	Sweep→retournement→confirmation, FVG post-sweep, retest
M30/M15/M5 Kijun	2	Kijun M30/M15/M5 cross directionnel

Score maximum: 109 critères (107 sans MN)

5. Pipeline Machine Learning

Le pipeline ML transforme les résultats du scanner en probabilité de gain via 3 étapes :

Étape	Outil	Détails
1	ml_labeler.py	Extraction des features depuis le scan Diamond (DB / CSV manuel / backtest). Génération du CSV labellisé avec colonne outcome (1=gagné, 0=perdu).
2	ml_trainer.py	Entraînement XGBoost avec cross-validation 5-fold, scale_pos_weight pour classes déséquilibrées, grid search hyperparamètres.
3	ml_predictor.py	Inférence temps réel : extract_features() → predict_proba() → ML% affiché dans le scan live. MLPredictor simplifié : 1 joblib.load au lieu de 5.

6. Modèle XGBoost v18 — Configuration

Le modèle v18 est le modèle de production actif. Il a été entraîné sur 1 210 trades historiques avec 123 features nommées (incluant GARCH et stochastique).

Hyperparamètre	Valeur
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 (basé sur ratio win/loss)

123 features nommées incluent : scores/biais (6), Kijun 6 TFs (18), Tenkan 6 TFs (18), T/K crosses (6), Kumo (6), M5/M15/M30 crosses (12), Flat bars (2), FVG actifs (1), Order Blocks (6), Sweeps (5), Volume (6), MSS/BOS (6), Breaker Blocks (6), Compression (1), Stochastique (8 : OU, Hurst, Monte Carlo, GARCH), Temporelles (3), Type d'actif (3).

7. Historique des Modèles

27 modèles XGBoost ont été entraînés et évalués depuis le début du projet. Voici l'historique des versions principales :

Modèle	Données	Features	Trades	CV	OOS PF
v1	6 majors	36	852	61.0%	—
v4	13 symboles	111	1 210	62.9%	1.26*
v7	13 symboles	25 best	2 923	60.1%	1.173
v13	4 modèles/actif	25	2 923	53-64%	4.000
v15	2025-2026	25 + OU/Hurst	5 009	59.8%	—
v18	13 symboles	123 (GARCH)	1 210	64.0%	0.503 F1

8. Importance des Features (v4)

L'analyse de feature importance révèle quels critères pèsent le plus dans les décisions du modèle :

Rang	Feature	Gain	Insight
1	is_dxy	14.2	Le dollar suit une logique radicalement différente des paires forex
2	rr	10.0	Le ratio risk/reward théorique prédit le résultat réel
3	tkx_h1	9.2	La croix Tenkan/Kijun H1 est le meilleur prédicteur unique
4	kj_h1	6.9	La distance à la Kijun H1 capture l'équilibre court-terme
5	s1	6.4	Le niveau de stop-loss est aussi important que le take-profit
6	day_wednesday	6.4	Les mercredis (FOMC) ont un comportement statistiquement distinct
7	d_kj_h4	6.3	La distance à la Kijun H4 indique la tendance de fond
8	d_kj_d1	6.0	La distance à la Kijun D1 ancre le contexte macro
9	tkx_h4	6.0	Le croisement T/K H4 confirme la tendance de session
10	is_forex	5.9	Les paires FX ont un comportement propre vs indices/métaux

9. Validation Out-of-Sample

La validation OOS (entraînement Janvier-Juin 2026, test Juillet 2026) est le test décisif. Le modèle v4 affiche un PF in-sample de 2.27 au seuil $ML\% \geq 50\%$, mais ce chiffre inclut du data leakage (le modèle a vu 80% des données de test pendant l'entraînement). Le vrai PF OOS du v7 est de 1.173 — un edge statistiquement significatif mais modeste. Le modèle v13 (per-asset) atteint PF=4.000 sur 15 trades, mais l'échantillon est trop petit pour être fiable. La conclusion : le ML apporte un edge réel ($PF > 1.0$), mais le plafond actuel est ~ 1.17 avec une architecture unifiée. L'amélioration viendra de plus de données et potentiellement de modèles spécialisés par type d'actif.

10. Simulation de Capital — Filtrage ML%

Simulation sur 1 210 trades historiques avec capital initial de 10 000 € et risque 1% par trade :

Seuil ML%	Trades	Win Rate	PF (RR)	P&L (10k€)
Aucun filtre	1 210	37.6%	1.02	+1 442 €
ML% $\geq$ 45%	661	54.3%	1.91 □	+27 542 €
ML% $\geq$ 50%	504	59.7%	2.27 □□	+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. Finance Quantitative & Calcul Stochastique

6 fonctions de finance quantitative ont été intégrées au scanner et au pipeline ML. Ces fonctions calculent des métriques de risque, de volatilité et de régime de marché :

Fonction	Objectif	Feature ML	Rang v18
Ornstein-Uhlenbeck	Détecte l'overextension du prix par rapport à sa moyenne long-terme. Score 0-100% : >80% = retournement probable.	ou_score	Non testé
Hurst Exponent	Classifie le régime de marché : $H < 0.45$ = RANGE (mean-reversion), $H > 0.55$ = TREND (momentum).	hurst_h	Non testé
GARCH(1,1)	Modélise la volatilité conditionnelle. 4 features : persistence, long_run_vol, half_life, forecast_vol_1.	garch_persistence	#14
Monte Carlo	Simule 10 000 trajectoires de prix. Calcule P(SL touché) et P(TP touché).	mc_p_sl, mc_p_tp	Non testé
Kelly Criterion	Dimensionnement optimal des positions par type d'actif. Gain=0.00 (redondant avec SL/TP déjà optimisés).	kelly_full	#124
FIMO Ruin	Probabilité de ruine d'un compte prop firm avec drawdown 10% sur 100 trades simulés.	ftmo_ruin_prob	Non testé

12. Kelly Criterion par Type d'Actif

Le Kelly Criterion utilise le win rate réel par type d'actif (issu du backtest) plutôt qu'un win_rate=0.40 uniforme. Résultats :

Type d'Actif	Win Rate	Kelly Full	Kelly Half
DXY	73%	46.0%	23.0%
XAUUSD	42%	13.0%	6.5%
Forex	31%	3.5%	1.8%
Indices	38%	7.0%	3.5%
Crypto	35%	2.5%	1.3%

13. MLPredictor Simplifié (19/07/2026)

Le MLPredictor chargeait précédemment le même modèle 5 fois (1× par méthode de prédiction). La simplification du 19/07/2026 a réduit cela à 1 seul `joblib.load()`. Résultat : 30 lignes de code supprimées, temps de chargement divisé par 4, et élimination de la `_ASSET_MODEL_MAP` redondante. Le modèle v18 est maintenant le modèle unifié unique pour tous les types d'actifs.

14. Corrections de Bugs — 19/07/2026

Une revue de code approfondie le 19/07/2026 a identifié et corrigé 4 bugs :

#	Gravité	Fichier	Description
1	CRITIQUE	diamond_scanner.py	Code mort après <code>return</code> dans <code>ml_available</code> — les symboles n'étaient jamais activés. <code>initialize()</code> ne définissait pas <code>_initialized = True</code> .
2	HAUTE	stochastic_analytics.py	GARCH : si aucune paire (α, β) valide trouvée, <code>fallback</code> silencieux avec paramètres inventés. Ajout d'une vérification <code>best_ll <= -1e100</code> .
3	MOYENNE	stochastic_analytics.py	Monte Carlo : le <code>fallback</code> de volatilité (<code>log-returns</code>) était calculé mais jamais utilisé quand le <code>sigma</code> OU était nul.
4	BASSE	stochastic_analytics.py	FTMO : <code>pnl_avg/roi_avg</code> hors du bloc <code>try</code> — risque de <code>NameError</code> si code ajouté entre le calcul et le <code>return</code> .

15. Matrice de Décision Finale

La matrice de décision combine le score Diamond, le ML%, les garde-fous et le contexte macro :

Condition	Action	Justification
ML% $\geq$ 60% + TKx HI aligné + pas HIGH NEWS	☐ GO	Triple confirmation : ML, Ichimoku, macro
ML% $\geq$ 60% + TKx HI conflit	☐ WAIT	Le ML est contredit par le croisement T/K HI
ML% 50-60% + TKx HI aligné	☐ ATTENTION	Edge ML modéré, nécessite confirmation additionnelle
ML% < 50%	☐ NO GO	Le modèle n'a pas d'edge statistique
HIGH NEWS DAY + BULL	☐ WAIT	Les jours macro (≥ 2 événements majeurs) cassent les corrélations
Mercredi (FOMC day) + Forex	☐ Risque élevé	Les mercredis ont un comportement statistiquement distinct

16. Conclusion et Perspectives

Le projet InelidaMarketScan démontre qu'un pipeline ML entraîné sur des features Ichimoku peut générer un edge statistique significatif (PF OOS = 1.173). Les 4 bugs critiques corrigés le 19/07/2026 renforcent la fiabilité du scanner live. Les 27 modèles successifs ont révélé des insights clés : les features Ichimoku (distances Kijun/Tenkan) sont plus prédictives que les patterns ICT, le type d'actif (is_dxy, is_forex) est un prédicteur majeur, et les fonctions de calcul stochastique (GARCH, OU, Hurst) apportent un signal complémentaire sur la volatilité. Le prochain saut de performance nécessitera un dataset plus large (2024-2026, >5000 trades/actif) et potentiellement des architectures deep learning (LSTM, Transformer) pour capturer les dépendances temporelles que XGBoost ne modélise pas.

Références

- 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, Transactions of the American Society of Civil Engineers, 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