Université Chouaib Doukkali
Faculté des Sciences d'el Jadida
Master 2IAD

جـامـعـة شعيـب الدكـالــي
Université Chouaib Doukkali

Project

ML4T (Machine Learning for Algorithmic Trading)

Adaptation des Stratégies ML au Marché des Crypto-monnaies

Réalisé par:

Hassan EL AZZOUZI

Yassine CHOUAYT

Encadré par:

Pr. A. BENIHSAINE

MASTER 2IAD - 2025/2026

I. INTRODUCTION

Ce rapport synthétise les travaux de recherche et de développement menés dans le cadre du projet 2IAD - ML4T. L'objectif principal était d'adapter les méthodologies de trading algorithmique, telles que définies par Stefan Jansen, aux spécificités du marché des crypto-monnaies.

Le marché des cryptoactifs présente des défis uniques par rapport aux marchés boursiers traditionnels : une ***volatilité extrême***, une ***cotation 24/7***, et une ***structure de frais élevée***. Ce document détaille notre parcours, de l'implémentation initiale des stratégies boursières classiques à leur évolution vers une architecture hybride "Fusion" intégrant le Traitement du Langage Naturel (NLP).

II. L'APPROCHE INITIALE: LA STRATÉGIE STEFAN JANSEN

Notre point de départ fut l'implémentation rigoureuse du workflow décrit par Stefan Jansen dans "Machine Learning for Algorithmic Trading". Cette approche, conçue pour les marchés actions, repose sur des principes fondamentaux que nous avons d'abord validés.

1. **Philosophie : "High Accuracy is a Bug"**
   Contrairement aux applications classiques du Machine Learning où l'on vise 90%+ de précision, en finance quantitative, un modèle affichant une telle précision est souvent victime de sur-apprentissage (overfitting).
   - **Cible Réaliste :** Nous avons visé une précision directionnelle de 52% à 54%.
   - **Métrique Clé :** L'accent a été mis sur le Coefficient d'Information (IC), ciblant un score de 0.05 à 0.10, mesurant la corrélation entre nos prédictions et les rendements futurs réels.
2. **Architecture Technique**
   Nous avons adopté la "World Class Stack" de Jansen pour les données tabulaires bruitées :
   - **Modèle Principal :** LightGBM (Gradient Boosting). Ce modèle excelle sur les données tabulaires financières où le rapport signal/bruit est faible, surpassant souvent les réseaux de neurones profonds (Deep Learning) sur ce type de données spécifiques.
   - **Ensembling (Stacking) :** Combinaison de Forêts Aléatoires (pour la non-linéarité), LightGBM (pour la gestion des gradients) et Régression Linéaire (pour ancrer les tendances).
   - **Validation :** Utilisation stricte de fenêtres glissantes (Walk-Forward Validation) pour éviter le biais de "Look-Ahead".

III. ADAPTATION AU MARCHÉ CRYPTO & RÉSULTATS

La transition des actions vers la Blockchain a nécessité une refonte majeure de notre pipeline de données et de nos caractéristiques (features).

1. **Les Défis de la Blockchain**
   - **Continuité Temporelle :** Absence de "clôture" quotidienne. Nous sommes passés à des fenêtres glissantes (Rolling Windows) de 5 et 15 minutes.
   - **Vélocité des Données:** Le flux de données WebSocket (Tick data) a imposé l'abandon des fichiers CSV au profit du format HDF5, permettant un stockage hiérarchique et une récupération ultra-rapide (<1.5s pour le pipeline complet).
2. **Nouveaux Facteurs "On-Chain"**
   Les ratios financiers classiques (P/E, EPS) n'existant pas en crypto, nous avons ingénieré des facteurs natifs à la blockchain :
   - **Flux des "Whales" (Baleines) :** Suivi des mouvements de portefeuilles importants comme indicateur avancé de liquidité.
   - **Volatilité du Gas (Ethereum) :** Utilisation des frais de transaction comme proxy de l'activité et de la congestion du réseau (souvent corrélée aux sommets de marché).
3. **Résultats Obtenus**
   L'adaptation de l'architecture XGBoost/LightGBM sur ces nouvelles données a produit des résultats supérieurs aux attentes initiales :
   - **Précision Backtestée :** Nos modèles optimisés ont atteint une précision de 64.2% à 70.1% sur des intervalles de 5 minutes durant les régimes de haute volatilité.
   - **Latence :** Le système complet (ingestion > inférence) opère en moins de 1.5 secondes, ce qui est acceptable pour du trading haute fréquence non-HFT.

IV. LE PROBLÈME DES FRAIS ("THE FEE TRAP")

Malgré une précision brute encourageante, nous avons identifié un obstacle critique : le "Piège des Frais".

1. **Analyse du Problème**
   Sur les marchés crypto, les frais de transaction (Maker/Taker) et le slippage (glissement de prix) sont nettement plus élevés que sur les marchés actions traditionnels.
   - Un modèle avec 55% de précision qui trade 100 fois par jour peut finir perdant si chaque trade coûte 0.1% en frais + slippage.
   - L'avantage statistique (Edge) est littéralement "mangé" par les frictions du marché.
2. **La Solution : Filtre de Confiance**
   Pour contrer cela, nous avons modifié notre logique d'exécution pour privilégier la Précision au détriment du Rappel (Recall).
   - **Seuil de Conviction :** Le système n'exécute un ordre que si la probabilité de confiance du modèle dépasse 53%.
   - **Résultat :** Cela réduit drastiquement le nombre de trades, mais augmente significativement l'espérance mathématique de chaque trade exécuté. Nous ne "tirons" que lorsque la cible est claire.

V. STRATÉGIE NLP (SOCIAL SENTIMENT INTEGRATION):

Pour franchir un nouveau palier de rentabilité, nous avons intégré une dimension souvent ignorée par l'analyse technique pure : le sentiment de marché.

1. **Pourquoi le NLP ?**
   Les crypto-monnaies sont des actifs hautement spéculatifs, souvent pilotés par la "Hype", les nouvelles réglementaires ou les tendances sociales (Twitter/X, Reddit). L'analyse technique seule (prix/volume) réagit souvent après que l'événement social ait eu lieu.
2. **Architecture "Dual-Model Fusion"**
   Nous avons développé une nouvelle approhe qui combine deux sources d'intelligence distinctes :
   - **Moteur Technique (60% du poids) :** Notre modèle XGBoost éprouvé (Prix, RSI, Bandes de Bollinger, Données On-Chain).
   - **Moteur de Sentiment (40% du poids) :** Un module NLP analysant :
     - ✓ **Score de Sentiment :** Positivité/Négativité des news.
     - ✓ **Volume Social :** Intensité des discussions.
     - ✓ **Polarité des News :** Impact potentiel des gros titres.
3. **Résultats de la Fusion**
   L'approche fusionnée permet de filtrer les faux signaux techniques. Par exemple, une rupture technique haussière (Breakout) non soutenue par un volume social est souvent un piège ("Bull Trap").
   - Le système actuel calcule une probabilité fusionnée.
   - Exemple de Logique : Si Technique = ACHAT (0.60) mais Sentiment = VENTE (0.40), le signal global est modéré, évitant une entrée risquée.

VI. CONCLUSION

Le projet 2IAD - ML4T a permis de démontrer que les stratégies de Machine Learning sont transposables au marché des crypto-monnaies, à condition de subir des adaptations structurelles majeures.

- L'infrastructure doit passer du CSV au HDF5/WebSocket pour gérer la vélocité.
- La stratégie d'exécution doit être défensive (Filtre de Confiance > 53%) pour survivre aux frais élevés.
- L'Alpha véritable réside dans l'hybridation des données: combiner la précision mathématique des indicateurs techniques avec la réactivité émotionnelle de l'analyse NLP.