



RAPPORT INTERMÉDIAIRE PSC PRÉDICTION DES TRENDS SUR LE MARCHÉ FOREX APRÈS DES MOUVEMENTS IMPORTANTS

24 janvier 2025

Abdelbar Ghassoub Bidias Cabrel Tiotso Ngueguim
Eliav Kabla Le Duc Wesnil Dzangueu-Djouda
Tuteur : Timon Fugier Fabrice Daniel

luis

Table des matières

1	INTRODUCTION	1
1.1	Présentation et motivations	1
1.2	État de l'art	1
1.3	Objectifs	2
2	Présentation des données	3
2.1	Description des données Forex	3
2.2	Enrichissement des données	3
3	Modèle de Classification - Random Forest	5
3.1	Introduction au Random Forest	5
3.2	Pourquoi utiliser le Random Forest dans le Forex ?	5
3.3	Comparaison avec d'autres modèles	6
3.4	Optimisation des hyperparamètres	6
4	Modèle de Classification - XGBoost	7
4.1	Introduction à XGBoost	7
4.2	Pourquoi utiliser XGBoost dans le Forex ?	7
4.3	Comparaison avec d'autres modèles	8
4.4	Optimisation des hyperparamètres	8
5	Synthèse des Résultats	9
5.1	Analyse Globale des Performances	9
5.2	Comparaison des Performances : Random Forest vs XGBoost	9
6	Avancement du projet	11
6.1	Difficultés Rencontrées	11
6.2	Opportunités Identifiées	11
6.3	Organisation et Collaboration	11
A	Algorithme : Détection et Analyse des Forts Mouvements	13
B	Résultats : Random Forest	14
B.1	Analyse des résultats sur différentes paires de devises	14
B.1.1	AUDUSD	14
B.1.2	EURJPY	14
B.1.3	EURUSD	15

B.1.4	GBPJPY	15
B.1.5	GBPUSD	16
B.1.6	NZDUSD	17
B.1.7	USDCAD	17
B.1.8	USDCHF	18

C	Résultats avec XGBoost	20
C.0.1	EUR/USD	20
C.0.2	AUD/USD	20
C.0.3	EUR/JPY	22
C.0.4	GBP/JPY	22
C.0.5	GBP/USD	23
C.0.6	NZD/USD	24
C.0.7	USD/CAD	24
C.0.8	USD/CHF	25
C.0.9	USD/JPY	25

Remerciements

Nous tenons d'abord à remercier chaleureusement **Julien COMBE**, notre coordinateur, pour la proposition de ce sujet, sa disponibilité et son aide dans le cadrage des attentes.

Nous remercions nos partenaires de **LUSIS**, sans qui ce projet n'aurait pu être mené à bout. Merci à **Fabrice DANIEL** pour la formation en finance de marché qu'il nous a dispensée, et pour l'indication de ressources utiles en intelligence artificielle.

Nous remercions grandement **Timon Fugier** pour son suivi hebdomadaire, son aide dans le guidage de nos recherches, et les nombreuses solutions qu'il nous a apportées et indiquées. Son aide a été capitale.

Enfin, nous remercions l'**École polytechnique**, et en particulier le **département d'économie**, dont le support logistique a permis l'établissement de ce travail.

Chapitre 1

INTRODUCTION

1.1 Présentation et motivations

Dans un environnement financier de plus en plus complexe et volatile, la capacité à anticiper les mouvements significatifs constitue un atout stratégique majeur. En particulier, sur le marché Forex, où les fluctuations sont fréquentes et influencées par de nombreux facteurs, être capable de prédire les tendances futures devient un avantage concurrentiel essentiel. Ce projet vise à utiliser les méthodes de Machine Learning pour détecter les événements horaires importants sur le marché Forex et prédire les mouvements à venir. L'objectif est de fournir des outils robustes permettant de prendre des décisions éclairées et optimales dans le cadre du trading algorithmique.

1.2 État de l'art

Pendant de nombreuses années, les analystes financiers se sont principalement appuyés sur des modèles statistiques classiques, tels que l'ARIMA (AutoRegressive Integrated Moving Average), pour analyser les séries temporelles. Ces outils, bien qu'efficaces pour identifier des tendances simples, montrent leurs limites face à des relations complexes ou des volumes importants de données multidimensionnelles.

Avec l'émergence du *Machine Learning*, de nouvelles approches ont vu le jour pour repousser ces limites. Par exemple, les réseaux de neurones récurrents, et plus précisément les modèles LSTM (*Long Short-Term Memory*), occupent aujourd'hui une place importante. Introduits dans les années 1990, ces modèles sont particulièrement performants pour détecter des dépendances à long terme dans des séries temporelles. Leur efficacité a déjà été démontrée dans des domaines exigeants, comme la prédiction des mouvements de marchés financiers ou l'identification d'anomalies.

Plus récemment, des approches hybrides combinant des réseaux neuronaux avec des mécanismes d'attention ont permis d'améliorer encore davantage les performances des modèles prédictifs. Grâce à ces mécanismes, les modèles peuvent se concentrer sur les données les plus pertinentes d'une séquence, ce qui ouvre de nouvelles perspectives dans l'analyse des marchés. Par ailleurs, des architectures comme les transformers, par exemple BERT ou GPT, initialement développées pour le traitement du langage, commencent également à faire leurs preuves dans le domaine des séries temporelles.

En parallèle, d'autres méthodes, basées sur des techniques d'ensemble comme le *bagging* ou le *boosting*, ont gagné en popularité. Ces approches reposent sur l'idée de combiner plusieurs modèles pour améliorer leur robustesse globale. Parmi les algorithmes populaires, on peut citer les *Gradient Boosting Machines* (GBM), AdaBoost ou encore CatBoost, qui ont montré une efficacité remarquable

pour traiter des problématiques liées aux données financières. Ces modèles, qui s'appuient sur des arbres de décision, sont capables de capturer des relations complexes entre différentes variables.

Enfin, deux méthodes sortent du lot grâce à leur grande efficacité et leur flexibilité : Random Forest et XGBoost. Ces modèles sont aujourd'hui largement reconnus par les experts pour leurs résultats dans l'analyse des données financières.

1.3 Objectifs

L'idée principale est de concevoir des modèles capables d'identifier automatiquement les fluctuations importantes sur une base horaire et de prévoir les tendances futures. Les objectifs principaux sont les suivants :

- **Identifier les fluctuations majeures** : Mettre en place des modèles capables de détecter, à l'échelle horaire, les événements significatifs qui influencent le marché Forex.
- **Prédire les tendances à venir** : Développer des outils permettant de prévoir les évolutions futures du marché, qu'elles soient à la hausse ou à la baisse.
- **Analyser les performances** : Évaluer l'efficacité des modèles grâce à des indicateurs précis et des tests rétrospectifs (backtesting) pour en valider la pertinence.

Chapitre 2

Présentation des données

2.1 Description des données Forex

Les données exploitées dans ce projet proviennent du marché des devises fourni par Lusis. Ces marchés, connus pour leur grande liquidité et leur fonctionnement en continu, 24 heures sur 24, constituent un terrain idéal pour l'analyse technique et la prévision des mouvements de devises. Les données historiques utilisées s'étendent sur une période allant de 2010 à 2021, avec une fréquence horaire. Elles incluent les valeurs OHLC (Open, High, Low, Close) pour chaque intervalle, offrant une vision détaillée des fluctuations au fil du temps. Ces séries temporelles servent de base essentielle pour analyser les tendances et développer des modèles prédictifs.

2.2 Enrichissement des données

Pour améliorer la précision des modèles prédictifs, plusieurs étapes d'enrichissement des données ont été mises en œuvre. Ces enrichissements incluent notamment le calcul d'indicateurs techniques classiques, la définition de seuils dynamiques et l'établissement d'une condition de «fort mouvement» basée sur ces indicateurs.

- **Calcul des indicateurs techniques** : Une série d'indicateurs techniques a été dérivée des données de prix (*OHLC*). Parmi ces indicateurs figurent [3] :
 - **ATR (Average True Range)** : mesure de la volatilité en évaluant l'amplitude des prix sur une période donnée.
 - **ADX (Average Directional Index)**, **+DI (Positive Directional Indicator)** et **-DI (Negative Directional Indicator)** : utilisés pour mesurer la force et la direction d'une tendance sur le marché.
 - **RSI (Relative Strength Index)** : un indicateur oscillant entre 0 et 100, permettant d'identifier les situations de surachat ou de survente.
 - **CMO (Chande Momentum Oscillator)** : un oscillateur qui mesure le momentum en fonction de la différence entre les gains et les pertes sur une période donnée.
 - **Oscillateur Stochastique (K et D)** : identifie les conditions de surachat ou de survente en se basant sur la position relative du prix de clôture par rapport à son amplitude sur une période donnée.
 - **ROC (Rate of Change)** : calcule le taux de variation des prix entre deux périodes pour évaluer le momentum.

- **EMA (Exponential Moving Average)** : des moyennes mobiles exponentielles calculées sur 10 et 20 périodes, permettant d’analyser les tendances de manière lissée.
- **Différences de prix (Price Differences)** : les variations de prix entre 1, 2 et 3 jours consécutifs, utilisées pour capturer les mouvements récents.
- **Identification des seuils dynamiques** : Pour chaque indicateur technique, des seuils dynamiques ont été définis à l’aide des percentiles. Par exemple, des seuils fixés aux 10% ou 20% des valeurs de chaque indicateur permettent d’établir des références utiles pour prendre des décisions de trading. Ces seuils aident à mieux comprendre les conditions de marché susceptibles de générer des mouvements significatifs.
- **Définition d’une condition de « fort mouvement »** : Une condition de « fort mouvement » sur le marché a été élaborée en combinant différentes métriques issues des indicateurs techniques. Par exemple, un «fort mouvement» peut être défini lorsque l’ADX dépasse un certain seuil, tandis que le RSI est supérieur à 70 (indiquant une situation de surchauffe du marché). Cette condition permet d’identifier des événements ayant une forte probabilité de provoquer des variations significatives sur le marché.

Données Initiales

Start date : 2010-01-03 23 :00 :00
End date : 2024-04-22 01 :00 :00

Date	Open	High	Low	Close	Label_UpBar
2010-01-03 23 :00 :00	1.43318	1.43364	1.42969	1.43246	0
2010-01-04 00 :00 :00	1.43246	1.43335	1.43148	1.43159	0
2010-01-04 01 :00 :00	1.43159	1.43197	1.42879	1.42885	0
2010-01-04 02 :00 :00	1.42885	1.42885	1.42569	1.42702	0
2010-01-04 03 :00 :00	1.42702	1.42995	1.42699	1.42942	1

Données après traitement :

date	open	high	low	close	...	ATR_condition	ADX_condition
2010-01-04 14 :00 :00	1.43961	1.44327	1.43869	1.44216	...	True	True
2010-01-04 15 :00 :00	1.44216	1.44498	1.44182	1.44211	...	True	True
2010-01-04 16 :00 :00	1.44211	1.44551	1.44027	1.44379	...	True	True
2010-01-04 17 :00 :00	1.44379	1.44408	1.44167	1.44173	...	True	True
2010-01-04 18 :00 :00	1.44173	1.44210	1.44120	1.44137	...	True	True

Chapitre 3

Modèle de Classification - Random Forest

3.1 Introduction au Random Forest

Le Random Forest est un algorithme d'apprentissage supervisé qui repose sur un ensemble d'arbres de décision. Conçu pour améliorer la précision et la robustesse des modèles prédictifs, il fonctionne en combinant plusieurs arbres, chacun étant construit à partir d'échantillons aléatoires des données et de sous-ensembles de caractéristiques.[1].

Principaux atouts et caractéristiques :

- **Résistance au surapprentissage** : En agrégeant les prédictions de plusieurs arbres, le Random Forest limite le risque de surapprentissage souvent rencontré avec un arbre de décision unique.
- **Robustesse face aux données bruitées** : Grâce à la technique du bootstrap, il devient moins sensible aux valeurs aberrantes et au bruit.
- **Capacité à gérer des données complexes** : Il est capable de modéliser efficacement les relations non linéaires et les interactions complexes entre les variables.
- **Facilité d'interprétation** : Des indicateurs comme l'importance des variables permettent d'identifier les caractéristiques les plus influentes dans le processus de décision.
- **Adaptabilité aux données déséquilibrées** : En ajustant les poids des classes ou en utilisant des techniques d'échantillonnage, il s'adapte aux jeux de données où certaines classes sont sous-représentées.

3.2 Pourquoi utiliser le Random Forest dans le Forex ?

Le marché des changes (Forex) est un environnement complexe, caractérisé par une forte volatilité, des dynamiques non linéaires et une influence de multiples facteurs, qu'ils soient exogènes ou endogènes. Le Random Forest s'avère être un outil particulièrement adapté pour relever ces défis, et ce, pour plusieurs raisons :

1. Capacité à gérer les relations non linéaires : Les fluctuations des devises résultent souvent d'interactions complexes entre divers indicateurs techniques et fondamentaux. Grâce à sa capacité à

capturer ces relations non linéaires, le Random Forest est particulièrement efficace pour modéliser ces dynamiques complexes.

2. Atténuation du bruit : Le marché Forex est sujet à un bruit important, notamment en raison d'événements imprévus tels que des annonces économiques soudaines. En combinant les prédictions de plusieurs arbres, le Random Forest permet de réduire l'impact de ces perturbations et d'améliorer la fiabilité des prévisions.

3. Identification des variables clés : Les traders et analystes cherchent souvent à identifier les indicateurs les plus influents sur les prédictions. Le Random Forest offre une mesure d'importance des caractéristiques, permettant de mieux comprendre les facteurs déterminants et d'optimiser les stratégies de trading.

4. Gestion des données déséquilibrées : Sur le marché des devises, les tendances marquées à la hausse ou à la baisse sont souvent moins fréquentes que les périodes de faible volatilité. Grâce à l'ajustement des poids des classes, le Random Forest est capable de traiter efficacement ce déséquilibre et d'améliorer la précision des modèles prédictifs.

3.3 Comparaison avec d'autres modèles

- **Régression logistique :** Bien qu'utile pour des relations linéaires simples, elle est souvent insuffisante pour capturer la complexité du Forex.
- **Support Vector Machines (SVM) :** Bien que performant pour des jeux de données réduits, il est moins efficace pour des données massives et bruitées comme celles du Forex.
- **Réseaux de neurones :** Bien qu'extrêmement puissants, ils nécessitent un grand volume de données et une optimisation approfondie, ce qui les rend parfois moins pratiques pour des projets avec des ressources limitées.

3.4 Optimisation des hyperparamètres

L'un des points forts du Random Forest est la personnalisation via des hyperparamètres tels que :

- **Nombre d'arbres (n_estimators) :** Augmenter ce paramètre améliore souvent la stabilité et la précision, au prix d'un temps de calcul accru.
- **Profondeur maximale (max_depth) :** Contrôle la complexité des arbres individuels pour éviter le surapprentissage.
- **Critère de séparation (gini, entropy) :** Permet de choisir la métrique utilisée pour mesurer la qualité des divisions.
- **Fraction d'échantillons (max_samples) :** Ajuste la proportion des données utilisées pour construire chaque arbre, influençant la diversité des arbres dans la forêt.

Chapitre 4

Modèle de Classification - XGBoost

4.1 Introduction à XGBoost

XGBoost (Extreme Gradient Boosting) est un algorithme d'apprentissage supervisé reposant sur la technique du gradient boosting. Il se démarque par sa rapidité, son efficacité et sa capacité à produire des modèles précis et optimisés, aussi bien pour les tâches de classification que de régression. [2].

Caractéristiques principales :

- **Excellente performance** : Grâce à l'optimisation des calculs et à une utilisation efficace des ressources, XGBoost est souvent considéré comme l'un des algorithmes les plus performants, en particulier dans les compétitions de machine learning.
- **Gestion intuitive des données manquantes** : L'algorithme traite automatiquement les valeurs manquantes en les intégrant dans une branche spécifique lors de la construction des arbres, ce qui simplifie la préparation des données.
- **Réduction efficace du surapprentissage** : L'intégration de mécanismes de régularisation permet de limiter le surapprentissage tout en conservant une précision élevée du modèle.
- **Grande flexibilité des fonctions de perte** : XGBoost prend en charge une large gamme de fonctions de perte, offrant ainsi une adaptabilité à divers types de problèmes et de besoins analytiques.
- **Prise en charge du multi-threading** : Grâce à une implémentation multi-thread, XGBoost garantit une exécution rapide, même sur de grands volumes de données.

4.2 Pourquoi utiliser XGBoost dans le Forex ?

Le marché des changes (Forex) est réputé pour sa dynamique constante, ses interactions complexes et la présence de signaux souvent subtils et bruités. XGBoost s'avère être un choix particulièrement pertinent pour relever ces défis, et ce, pour plusieurs raisons :

1. Modélisation des dynamiques complexes : Grâce à son approche basée sur des arbres de décision successifs et pondérés, XGBoost est capable de capturer des relations complexes entre les différentes variables, ce qui est essentiel dans un environnement hautement interactif comme le Forex.

2. Résistance au bruit des données : En utilisant une pondération adaptative des erreurs et des techniques de régularisation, XGBoost permet de minimiser l'impact des anomalies et des variations imprévisibles, rendant les prévisions plus fiables.

3. Identification des facteurs clés : L'algorithme fournit des indicateurs précis sur l'importance des variables, aidant ainsi les analystes à mieux comprendre quels facteurs influencent le plus les fluctuations des devises.

4. Rapidité et efficacité computationnelle : Grâce à sa capacité à traiter efficacement de vastes ensembles de données, XGBoost est idéal pour analyser les flux massifs d'informations générés en continu par le marché Forex.

5. Grande flexibilité d'adaptation : Avec ses nombreux hyperparamètres ajustables, XGBoost permet d'adapter finement le modèle aux particularités du marché des changes, offrant ainsi des prédictions plus pertinentes et adaptées aux besoins spécifiques des traders.

4.3 Comparaison avec d'autres modèles

- **Random Forest :** Bien que robustes, les forêts aléatoires peuvent manquer de précision dans des cas où une pondération adaptative des erreurs est cruciale.
- **Réseaux de neurones :** Bien qu'efficaces pour des données très complexes, ils demandent souvent plus de temps et de données pour être optimisés.
- **Gradient Boosting classique :** XGBoost est une version optimisée et plus rapide du gradient boosting, le rendant plus adapté à des problèmes avec des contraintes de temps.

4.4 Optimisation des hyperparamètres

Comme tout algorithme avancé, XGBoost peut être personnalisé pour améliorer ses performances. Voici quelques hyperparamètres clés :

- **Nombre d'arbres (`n_estimators`) :** Définit le nombre total d'arbres à construire. Une valeur élevée peut améliorer la précision, mais au prix d'un temps de calcul accru.
- **Taux d'apprentissage (`learning_rate`) :** Détermine la contribution de chaque arbre au modèle final. Des valeurs faibles (par exemple, 0.01 ou 0.1) réduisent le risque de surapprentissage.
- **Profondeur maximale (`max_depth`) :** Contrôle la complexité des arbres individuels. Une profondeur trop grande peut entraîner un surapprentissage, tandis qu'une profondeur trop faible peut limiter la performance.
- **Sous-échantillonnage (`subsample`) :** Définit la proportion des données utilisées pour chaque arbre. Cela réduit la variance et améliore la robustesse du modèle.
- **Colsample_bytree :** Indique la fraction des caractéristiques utilisées pour construire chaque arbre. Cela favorise la diversité dans les arbres et aide à éviter le surapprentissage.

Chapitre 5

Synthèse des Résultats

5.1 Analyse Globale des Performances

Les résultats montrent des performances variables en fonction des paires de devises et des métriques utilisées. De manière générale, les observations suivantes peuvent être faites :

Rentabilité : Les modèles ont affiché une rentabilité modérée à élevée sur certaines paires comme USD/CAD et NZD/USD, avec des *Profit Factors* supérieurs à 1.2, suggérant un potentiel d'exploitation intéressant.

Risque : Le *Max Drawdown* est un facteur clé, en particulier pour les paires à forte volatilité comme EUR/JPY et GBP/JPY, où les pertes maximales observées restent significatives, nécessitant une gestion stricte du risque.

Stabilité des Prédictions : L'*accuracy* des modèles oscille entre 51 % et 53 %, traduisant une capacité de prédiction légèrement supérieure au hasard, ce qui montre un besoin d'optimisation pour renforcer la robustesse des modèles.

Ratio Risque/Rendement : Les modèles présentent souvent des ratios de Sharpe faibles (inférieurs à 1), révélant un niveau de risque élevé par rapport aux rendements espérés. Cependant, certaines paires comme USD/CAD et NZD/USD affichent des ratios plus intéressants.

5.2 Comparaison des Performances : Random Forest vs XGBoost

Rentabilité (Profit Factor) XGBoost a démontré de meilleures performances en termes de rentabilité sur la majorité des paires, avec des *Profit Factors* plus élevés. Par exemple, pour la paire EUR/USD, XGBoost atteint un *Profit Factor* de 1.28 contre 1.03 pour Random Forest, montrant une meilleure gestion du rendement par rapport aux pertes.

Précision (Accuracy) : XGBoost affiche des taux de précision légèrement supérieurs à ceux de Random Forest, avec des moyennes atteignant jusqu'à 57 % pour certaines paires, contre environ 53 % pour Random Forest. Cette différence reflète une capacité de généralisation accrue de XGBoost.

Gestion du Risque (Sharpe Ratio et Max Drawdown)

Bien que XGBoost offre de meilleures performances en rentabilité, il est souvent associé à des risques plus élevés, comme l'indique le *Max Drawdown*, qui est parfois supérieur à celui de Random Forest. À l'inverse, Random Forest tend à proposer des solutions plus conservatrices, avec des *Sharpe Ratios* généralement plus faibles mais plus stables. Par exemple, pour la paire GBP/JPY, Random Forest affiche un drawdown maximal de \$127,210 contre \$135,900 pour XGBoost.

Stabilité des Performances

Random Forest montre une stabilité accrue sur les paires de devises à forte volatilité, avec des pertes moins sévères. En revanche, XGBoost est plus sensible aux fluctuations du marché, nécessitant un ajustement précis des hyperparamètres pour éviter le surajustement.

TABLE 5.1 – Comparaison des performances entre Random Forest et XGBoost

Paire	Profit Factor		Sharpe Ratio		Accuracy (%)	
	RF	XGB	RF	XGB	RF	XGB
EUR/USD	1.03	1.28	0.49	0.66	51.0	56.9
AUD/USD	1.02	1.10	0.44	0.22	53.0	55.3
GBP/USD	0.94	1.06	0.98	0.16	51.7	51.1
USD/CAD	1.16	1.34	0.68	1.06	51.0	45.7
NZD/USD	1.25	0.80	1.16	-0.54	52.0	50.5
USD/JPY	1.19	1.19	0.16	0.63	49.7	49.7

Chapitre 6

Avancement du projet

6.1 Difficultés Rencontrées

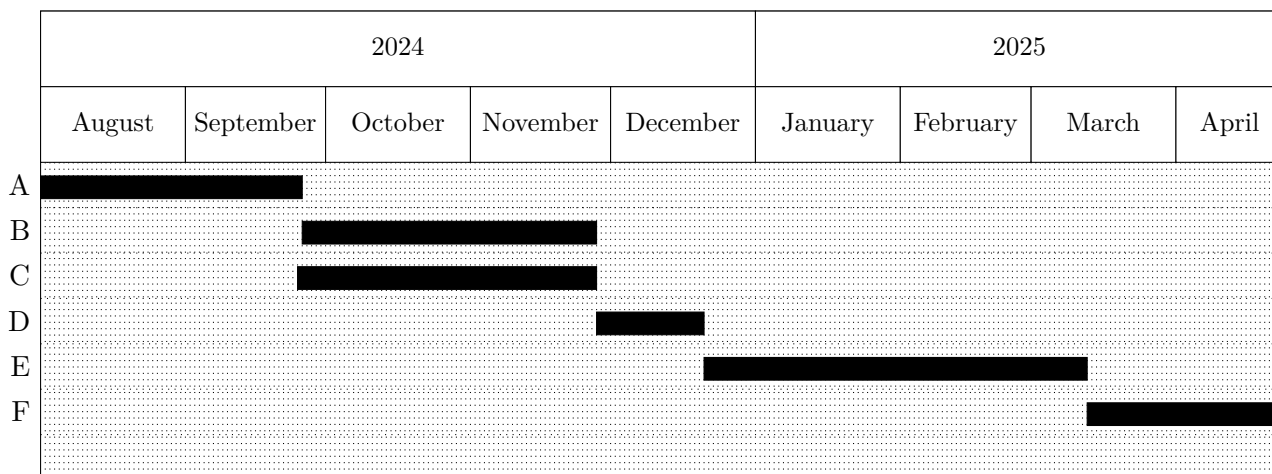
- **Compréhension du sujet** : Le marché Forex étant particulièrement complexe, il a fallu du temps pour maîtriser les concepts de base et avancer vers une modélisation précise.
- **Manipulation des librairies** : L'utilisation de librairies complexes comme Talib et AIutils a nécessité une phase d'apprentissage importante pour exploiter pleinement leurs fonctionnalités.
- **Fuite de données (data leaks)** : L'un des défis majeurs a été de s'assurer que les données utilisées pour l'entraînement des modèles ne contaminent pas les ensembles de validation ou de test, ce qui aurait pu biaiser les résultats.

6.2 Opportunités Identifiées

Exploration d'autres horizons temporels : Tester des prédictions à des horizons plus longs ou intermédiaires (par exemple, 1 jour ou 4 heures) pourrait offrir des insights supplémentaires.

Sélection avancée des caractéristiques : Une analyse approfondie pour identifier les indicateurs les plus pertinents permettrait d'améliorer la qualité des modèles et de réduire le bruit.

6.3 Organisation et Collaboration



Clé :

- **A_Tous** : Familiarisation avec les outils de base
- **B_Briyone** : Implémentation d'un algorithme de détection de forts mouvements horaires sur les paires de devises
- **C_Cabrel, Abdelbar, Eliav, Le Duc** : Développement de quelques modèles de classification :
 - Cabrel et Abdelbar pour le random-forest
 - Le Duc et Eliav pour le xgboost
- **D_Tous** : Test et backtesting des modèles sur les données historiques fournies par Lusiis
- **E_Tous** : Optimisation des modèles
- **F_Tous** : Rédaction du rapport final et des résultats

Références

1. StatQuest : Random Forests Part 1 - Building, Using and Evaluating https://www.youtube.com/watch?v=J4Wdy0Wc_xQ&t=154s.
2. XGBoost Partie 1 (sur 4) : Régression <https://www.youtube.com/watch?v=0tD8wVaFm6E>.
3. Lien, K. (2008). *Day Trading and Swing Trading the Currency Market : Technical and Fundamental Strategies to Profit from Market Moves*. Wiley Trading, 2nd Edition.

Annexe A

Algorithme : Détection et Analyse des Forts Mouvements

Algorithm 1 Détection des forts mouvements dans les séries temporelles

Require: Données historiques *df*, horizon *horizon*, lookback *lookback*, et caractéristiques *features*.

- 1: Charger les dépendances nécessaires : `pandas`, `numpy`, `sklearn`, `tensorflow`, `talib`, et `aiutils`.
- 2: Définir les paramètres globaux, incluant `symbol`, `lookback`, et `horizon`.
- 3: Créer les caractéristiques de volatilité, tendance, momentum et oscillateurs .
- 4: Charger les données avec `aiutils.read_timeseries_from_parquet`.
- 5: Ajouter des colonnes pour la détection des conditions :
 - Volatilité : Calculer `ATR_condition`.
 - Tendance : Calculer `ADX_condition`.
 - Momentum : Calculer `RSI_condition` et `CMO_condition`.
 - Oscillateurs : Calculer `Stochastic_condition` et `ROC_condition`.
- 6: Combiner les conditions pour identifier les "forts mouvements" (`strong_movement`).
- 7: Filtrer les données pour inclure uniquement les forts mouvements.
- 8: Ajouter des labels pour la classification, en utilisant `aiutils.labels_manager`.
- 9: Préparer les données pour l'entraînement (`pr.prepare_data`).
- 10: Définir les hyperparamètres du modèle `RandomForestClassifier`.
- 11: Entraîner le modèle sur le jeu d'entraînement.
- 12: Evaluer sur les données de validation pour savoir comment il se généralise.
- 13: Optimiser les hyperparamètres avec `RandomizedSearchCV` sur le jeu d'entraînement en observant s'il se généralise bien sur le jeu de validation.
- 14: Évaluer les performances du modèle sur les données de test (`confusion_matrix`, `accuracy_score`).
- 15: Effectuer un backtest avec les prédictions données du backtest (`aiutils.ai_backtester`).

Annexe B

Résultats : Random Forest

B.1 Analyse des résultats sur différentes paires de devises

B.1.1 AUDUSD

- **Profit Factor** : 1.02 (slightly profitable)
- **Sharpe Ratio** : 0.4462 (low but positive)
- **Max Drawdown** : \$592.3(3%)
- **Cumulative Return** : 3.82%
- **Accuracy** : 53%
- **Average Trade Win/Loss (ATWL)** : Global : 0,07 Long : 0.18; Short -0.01

Observation : Le modèle génère un gain total de 3,82% avec une *accuracy* de 53%, ce qui reste prévisible. Le ratio de Sharpe est de 0,45, tandis que le *max drawdown* atteint 592 \$. On remarque que les trades sur les positions longues rapportent en moyenne 18 cents par trade, alors que les positions courtes entraînent des pertes en moyenne. De plus, cet ATWL est insuffisant pour compenser les frais de transaction, rendant le modèle non rentable dans son état actuel. Cependant, il pourrait être amélioré en définissant un seuil de gain avant de sortir du marché. Une autre piste d'amélioration serait de se concentrer sur l'analyse des prédictions incorrectes afin d'optimiser les performances globales.

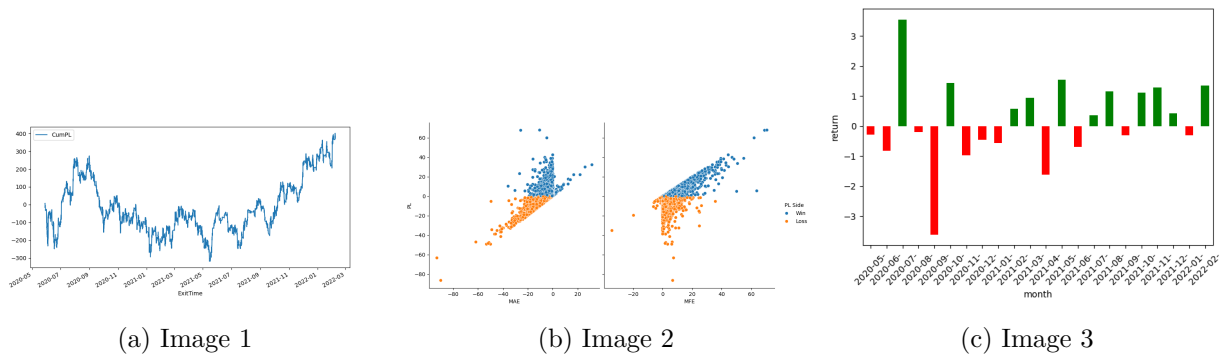


FIGURE B.1 – Résultats pour AUDUSD

B.1.2 EURJPY

- **Profit Factor** : 1.04 (marginally profitable)

- **Sharpe Ratio** : 0.8554(GOOD)
- **Max Drawdown** : \$70030
- **Cumulative Return** : 8.871 (très élevé)
- **Accuracy** : 53%
- **Average Trade Win/Loss (ATWL)** : Global : 16.11 Long : 67.66; Short -27.88

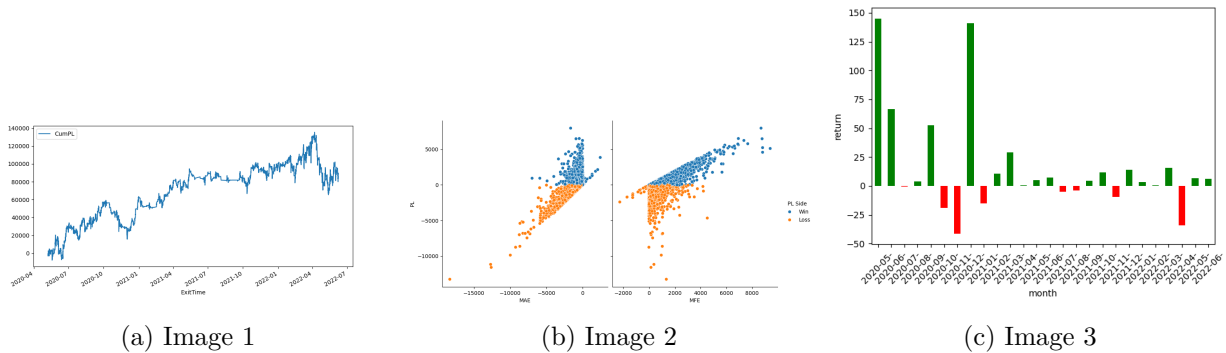


FIGURE B.2 – Résultats pour EURJPY

Observation : Les retours totaux du modèle sont exceptionnels, mais les *drawdowns* sont également très élevés. Avec une *accuracy* de 53,19%, le modèle génère un gain de 8,88%, ce qui est surprenant. Cela pourrait s'expliquer par la forte volatilité du taux de change entre l'euro et le yen : lorsqu'on gagne, les gains sont conséquents, mais les pertes sont également importantes. Par ailleurs, l'ATWL pour les positions longues est d'environ 67,66 yens, ce qui pourrait permettre de couvrir les frais de transaction. Cependant, en raison de cette volatilité, le modèle doit être utilisé avec précaution. Il serait donc pertinent d'apporter des améliorations au modèle afin de trader de manière plus prudente.

B.1.3 EURUSD

- **Profit Factor** : 1.03 (moderately profitable)
- **Sharpe Ratio** : 0.4902 (low but positive)
- **Max Drawdown** : 718
- **Cumulative Return** : 5.54%
- **Accuracy** : 51%
- **Average Trade Win/Loss (ATWL)** : Global : 0.10 Long : -0.08; Short 0.22

Observation : Le modèle présente une bonne performance malgré une *accuracy* de 51% et des *drawdowns* relativement faibles. Les ATWL restent cependant très faibles. Un point notable, qui diffère des autres modèles, est que l'on réalise des gains en *shortant* la monnaie, c'est-à-dire en pariant sur sa baisse. Toutefois, les retours moyens par trade sont faibles et ne permettent pas de couvrir les frais de transaction, estimés à environ 50 centimes par trade. Néanmoins, au vu des autres métriques du modèle, comme le ratio de Sharpe et le *max drawdown*, ce modèle montre un certain potentiel. Il serait pertinent d'explorer des améliorations avant d'envisager son déploiement.

B.1.4 GBPJPY

- **Profit Factor** : 1.04 (profitable)
- **Sharpe Ratio** : 0.3069(low)
- **Max Drawdown** : \$127,210

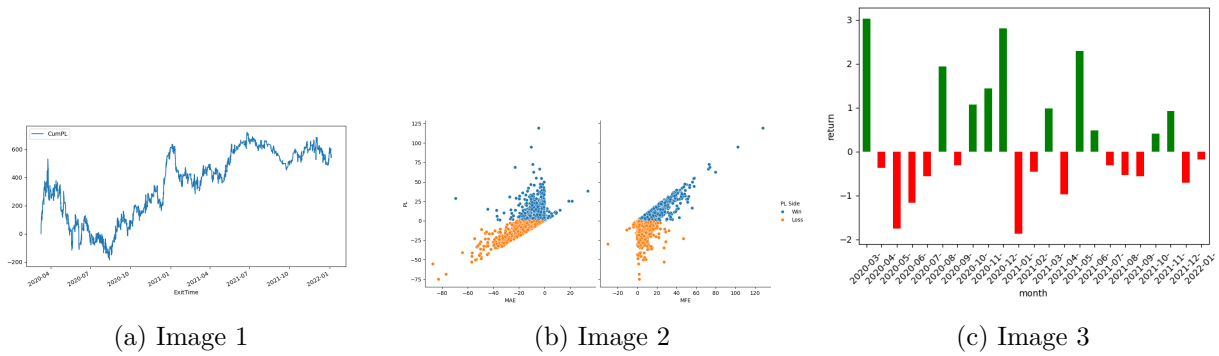


FIGURE B.3 – Résultats pour EURUSD

- **Cumulative Return** : 12.96%
- **Accuracy** : 52%
- **Average Trade Win/Loss (ATWL)** : Global : 23.53 Long : 60.00; Short -3.34

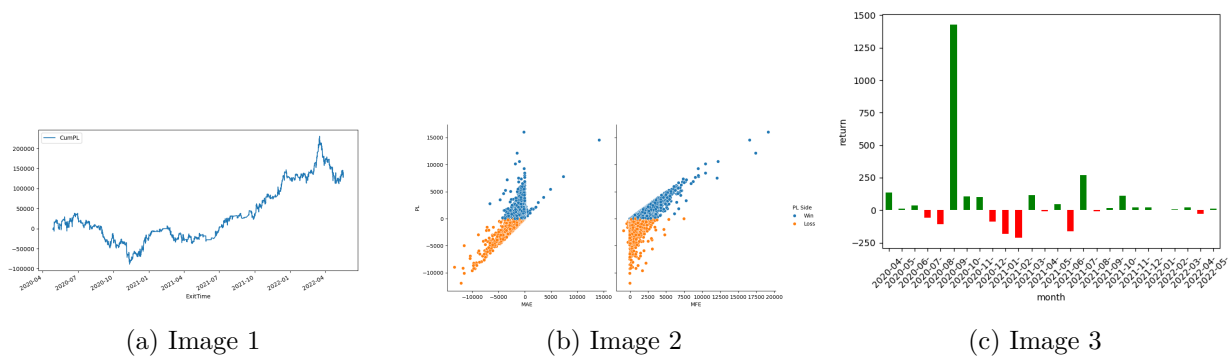


FIGURE B.4 – Résultats pour GBPJPY

Observation : Bien que le modèle génère un rendement de 12,96%, son *accuracy* est seulement de 52%. De plus, le ratio de Sharpe s'élève à 0,31, indiquant un niveau de risque relativement élevé. Ce rendement est principalement dû à des performances exceptionnelles enregistrées en août 2020, mais rien n'assure la reproductibilité de ces résultats. Par ailleurs, l'ATWL du modèle est très faible, ce qui limite considérablement son utilité pratique. En effet, en moyenne, le modèle ne parvient même pas à couvrir les coûts de transaction.

B.1.5 GBPUSD

- **Profit Factor** : 0.94
- **Sharpe Ratio** : 0.9873
- **Max Drawdown** : \$936.30 (9%)
- **Cumulative Return** : 0.14%
- **Accuracy** : 51,7%
- **Average Trade Win/Loss (ATWL)** : Global : 0.25 Long : 0.18; Short 0.31

Observation : Le modèle obtenu n'est pas très performant et engendre des pertes, bien que celles-ci restent modérées. Obtenir des gains aussi faibles avec ce modèle était prévisible, étant donné une

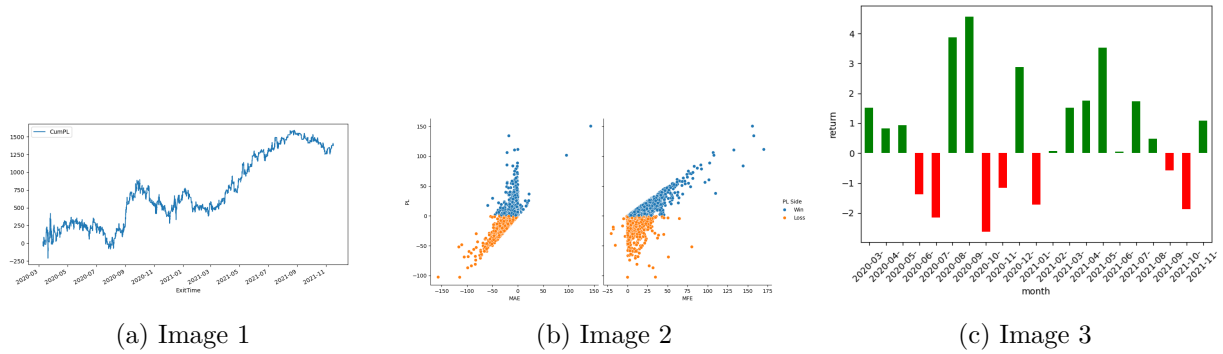


FIGURE B.5 – Résultats pour GBPUSD

accuracy de 51,7 %. Ce modèle présente un ratio de Sharpe d'environ 1, ce qui indique un faible niveau de risque et le rend potentiellement déployable, malgré des retours très faibles et des *Average Trade Win/Loss (ATWL)* bas.

B.1.6 NZDUSD

- **Profit Factor** : 1.25 (très profitable)
- **Sharpe Ratio** : 1.1670
- **Max Drawdown** : \$240.60 (2%)
- **Cumulative Return** : 11.45%
- **Accuracy** : 52%
- **Average Trade Win/Loss (ATWL)** : Global : 0.19 Long : 0.21; Short 0.18

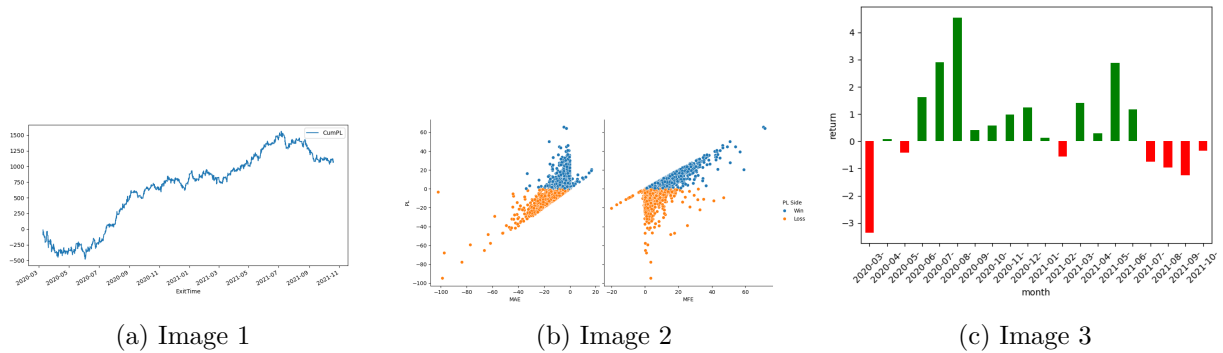


FIGURE B.6 – Résultats pour NZDUSD

Observation : Cette stratégie affiche de bonnes performances en termes de métriques, et les retours sont très encourageants. Une éventuelle mise en œuvre pourrait être envisagée. Cependant, il convient de rester attentif à l'*Average Trade Win/Loss (ATWL)*, qui reste relativement faible, même pour les positions longues, ce qui augmente le niveau de risque du modèle.

B.1.7 USDCAD

- **Profit Factor** : 1.16 (profitable)
- **Sharpe Ratio** : 0.68

- **Max Drawdown** : 731.7 (7%)
- **Cumulative Return** : 12.90%
- **Accuracy** : 51%
- **Average Trade Win/Loss (ATWL)** : Global : 0.14 Long : 0.09; Short 0.18

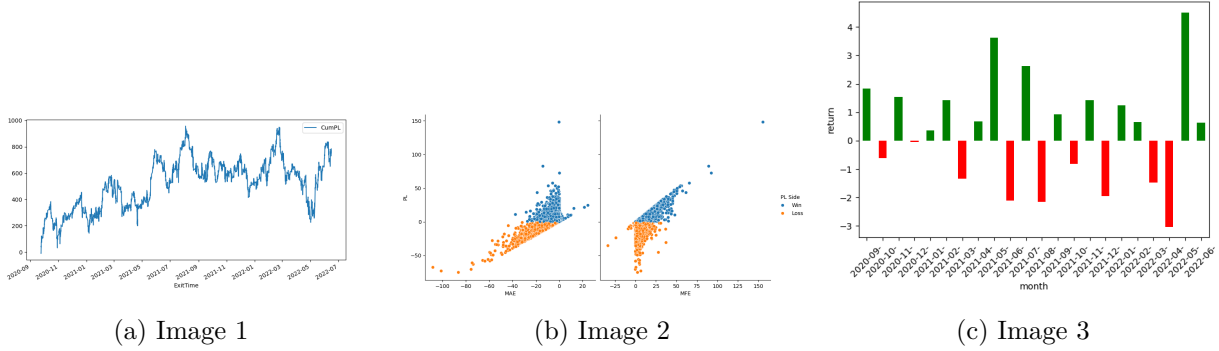


FIGURE B.7 – Résultats pour USDCAD

Observation : La stratégie est profitable et présente un bon ratio risque-rendement, ce qui lui confère une apparente robustesse. Les gains et les pertes sont réalistes. Le *max drawdown* est faible et le ratio de Sharpe est supérieur à 0.4, ce qui indique un faible niveau de risque. Avec ce modèle, en moyenne, les pertes sont évitées quelle que soit la position choisie. Cependant, les gains moyens par trade restent modestes.

B.1.8 USDCHEF

- **Profit Factor** : 1.16 (profitable)
- **Sharpe Ratio** : 1.6847
- **Max Drawdown** : 288.7 (3%)
- **Cumulative Return** : 0.14%
- **Accuracy** : 53%
- **Average Trade Win/Loss (ATWL)** : Global : 0.25 Long : 0.20; Short 0.30

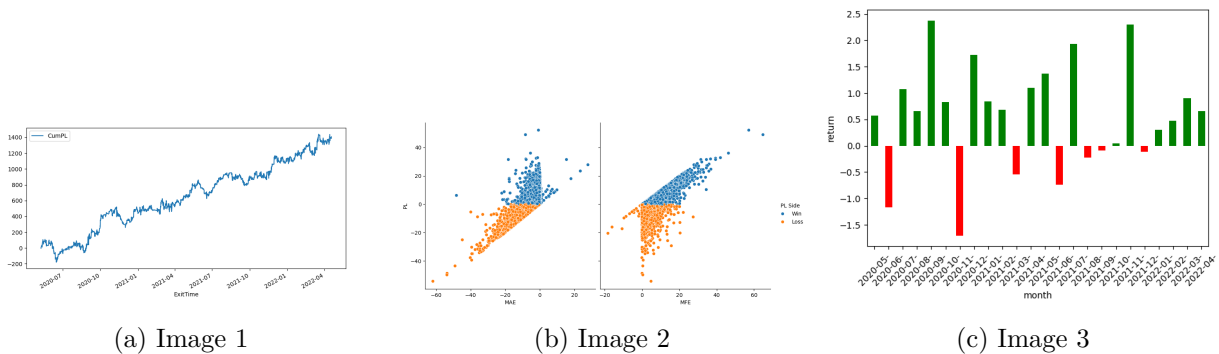


FIGURE B.8 – Résultats pour USDCHEF

Observation : La stratégie est profitable et présente un ratio risque-rendement intéressant, ce qui suggère une certaine robustesse. Cependant, les gains moyens par trade restent très faibles, au point

de ne pas suffire à couvrir les frais de transaction.

Analyse et Observations

Le modèle le plus performant en termes de retours a été obtenu sur la paire **USD/CAD**. Cette performance semble refléter une forte corrélation entre le dollar américain et le dollar canadien. Toutefois, en termes de ratio retour/risque, le modèle le plus efficace est celui appliqué à la paire **NZD/USD**. Bien que ce dernier présente de bons résultats dans ce contexte, les retours globaux demeurent modestes, comme en témoignent les courbes obtenues.

Annexe C

Résultats avec XGBoost

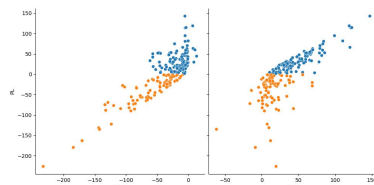
Analyse des résultats sur différentes paires de devises

C.0.1 EUR/USD

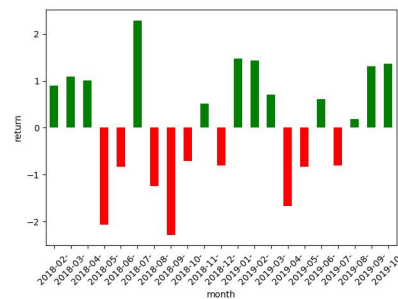
- Profit Factor : 1.28
- Sharpe Ratio : 0.6635
- Sortino Ratio : 0.9363
- AVG Win/Loss : 4.60
- Accuracy : 0.5698



(a) Equity



(b) P/L vs MAE et MFE



(c) Retours mensuels

FIGURE C.1 – Résultats pour EUR/USD

Observation : Les résultats sont encourageants, l'accuracy et le Profit factor sont plutôt bons, seul le sharpe semble un peu élevé, indiquant un niveau de risque élevé par rapport aux gains..

C.0.2 AUD/USD

- Profit Factor : 1.10
- Sharpe Ratio : 0.2273
- Sortino Ratio : 0.3146
- AVG Win/Loss : 1.74
- Accuracy : 0.5534

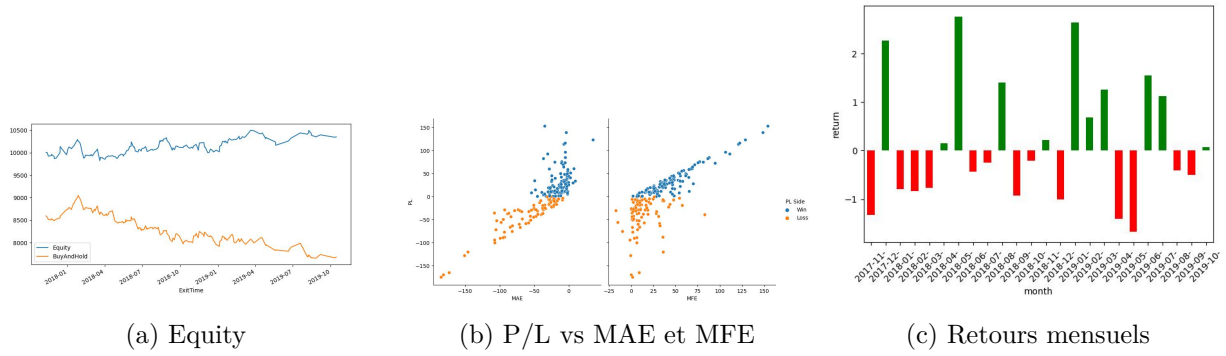


FIGURE C.2 – Résultats pour AUD/USD

Observation : Le modèle présente une performance acceptable avec un profit factor de 1.10 et une accuracy légèrement supérieure à 55%. Toutefois, le faible Sharpe Ratio (0.22) indique aussi un niveau de risque élevé par rapport aux gains.

C.0.3 EUR/JPY

- Profit Factor : 0.96
- Sharpe Ratio : NaN
- Sortino Ratio : NaN
- AVG Win/Loss : -105.75
- Accuracy : 0.5907

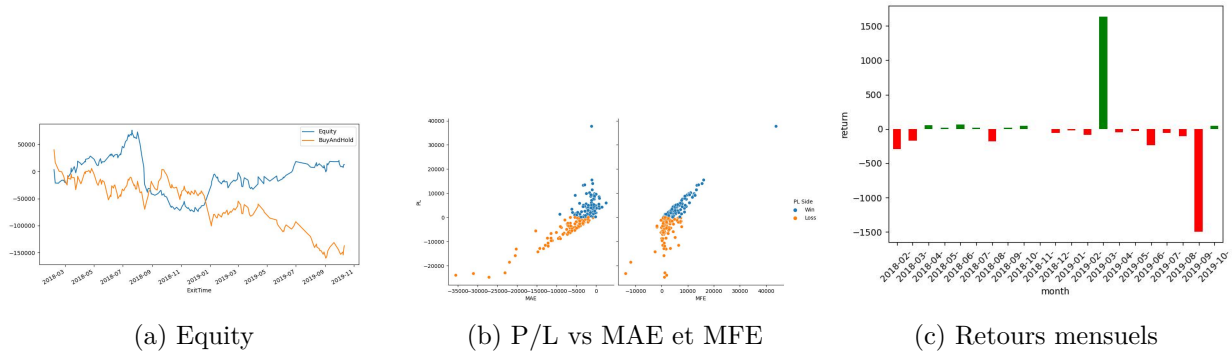


FIGURE C.3 – Résultats pour EUR/JPY

Observation : Malgré une accuracy raisonnable de 59%, le profit factor inférieur à 1 et le ratio AVG Win/Loss fortement négatif (-105.75) montrent que cette stratégie n'est pas profitable.

C.0.4 GBP/JPY

- Profit Factor : 0.86
- Sharpe Ratio : NaN
- Sortino Ratio : NaN
- AVG Win/Loss : -597.54
- Accuracy : 0.5670

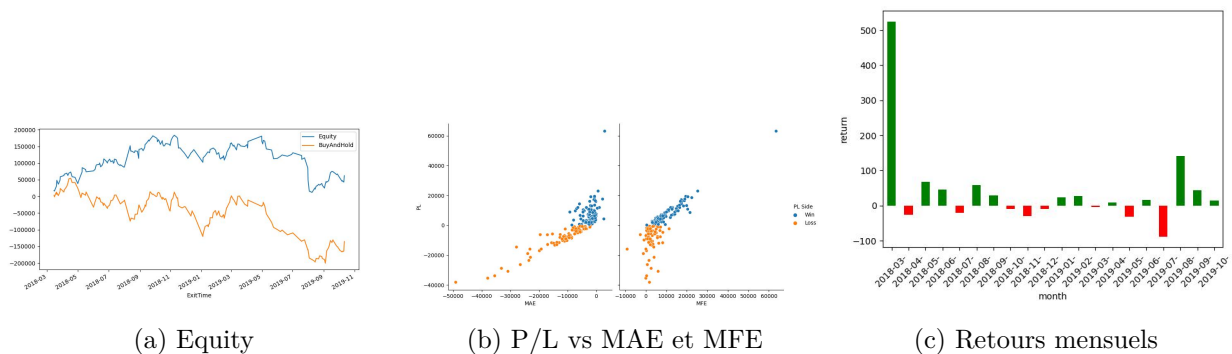


FIGURE C.4 – Résultats pour GBP/JPY

Observation : Bien que l'accuracy soit correcte, les résultats globaux sont décevants avec un profit factor inférieur à 1 et des pertes importantes.

C.0.5 GBP/USD

- Profit Factor : 1.06
- Sharpe Ratio : 0.1684
- Sortino Ratio : 0.2272
- AVG Win/Loss : 1.51
- Accuracy : 0.5114

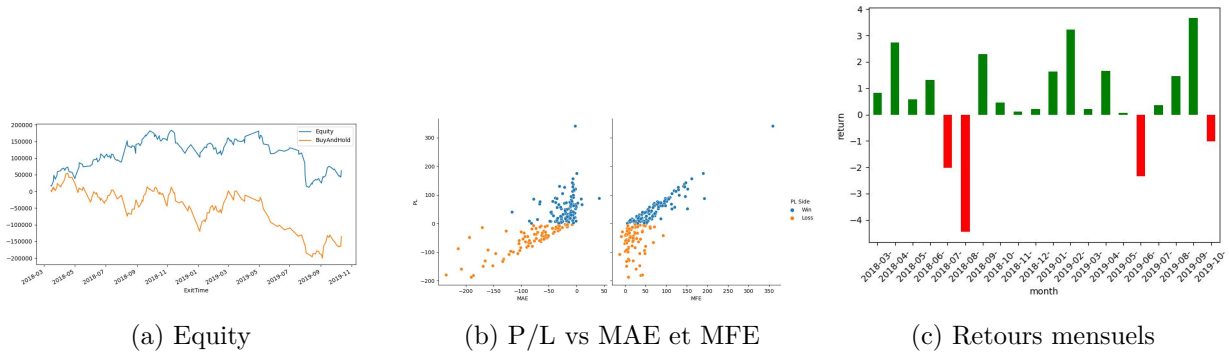


FIGURE C.5 – Résultats pour GBP/USD

Observation : Le modèle offre une performance équilibrée mais peu impressionnante avec un profit factor proche de 1 et un Sharpe Ratio faible (0.16). La faible AVG Win/Loss (1.51) indique que les gains sont modestes.

C.0.6 NZD/USD

- Profit Factor : 0.80
- Sharpe Ratio : -0.5499
- Sortino Ratio : -0.6999
- AVG Win/Loss : -3.90
- Accuracy : 0.5053

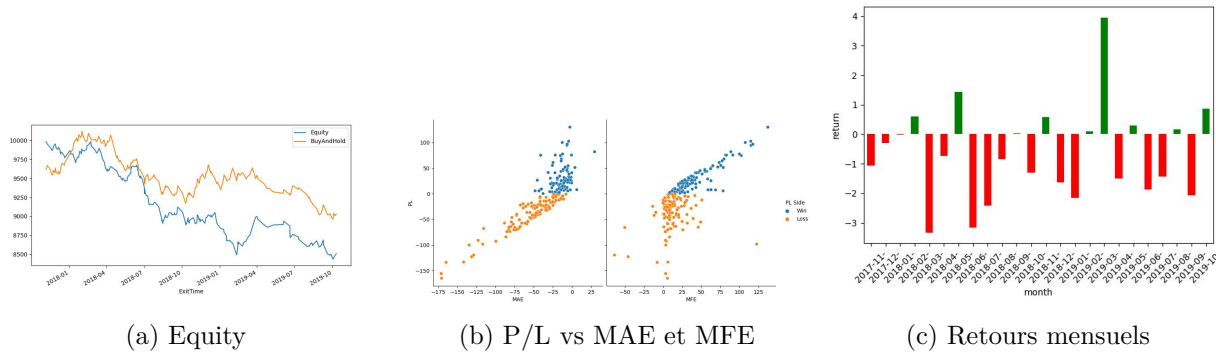


FIGURE C.6 – Résultats pour NZD/USD

Observation : La stratégie montre des résultats largement négatifs avec un profit factor inférieur à 1 et des ratios Sharpe et Sortino négatifs. L'AVG Win/Loss (-3.90) et l'accuracy de 50.53% confirment que ce modèle n'est pas viable dans sa configuration actuelle.

C.0.7 USD/CAD

- Profit Factor : 1.34
- Sharpe Ratio : 1.0596
- Sortino Ratio : 1.7186
- AVG Win/Loss : 5.63
- Accuracy : 0.4569

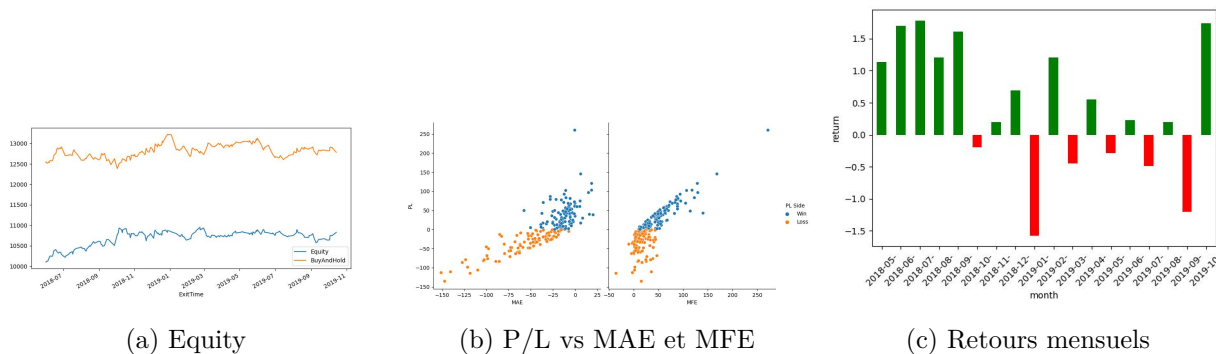


FIGURE C.7 – Résultats pour USD/CAD

Observation : Malgré une faible accuracy, le modèle se distingue par un profit factor solide (1.34) et des ratios Sharpe et Sortino élevés malgré une accuracy inférieure à 50%, reflétant un bon équilibre risque/retour. Cette stratégie est prometteuse et pourrait être déployée après validation.

C.0.8 USD/CHF

- Profit Factor : 1.05
- Sharpe Ratio : 0.1503
- Sortino Ratio : 0.2094
- AVG Win/Loss : 0.79
- Accuracy : 0.5317

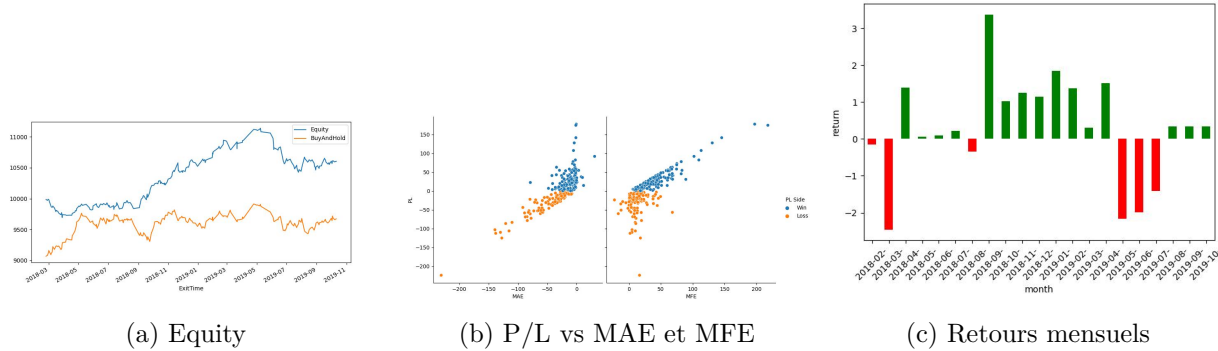


FIGURE C.8 – Résultats pour USD/CHF

Observation : La performance du modèle est très modérée avec un profit factor juste au-dessus de 1 et un Sharpe Ratio faible. Les gains modestes (AVG Win/Loss de 0.79) limitent l'intérêt de cette stratégie. Elle nécessite des ajustements pour améliorer sa rentabilité.

C.0.9 USD/JPY

- Profit Factor : 1.19
- Sharpe Ratio : 0.1681
- Sortino Ratio : 0.6343
- AVG Win/Loss : 275.69
- Accuracy : 0.4973

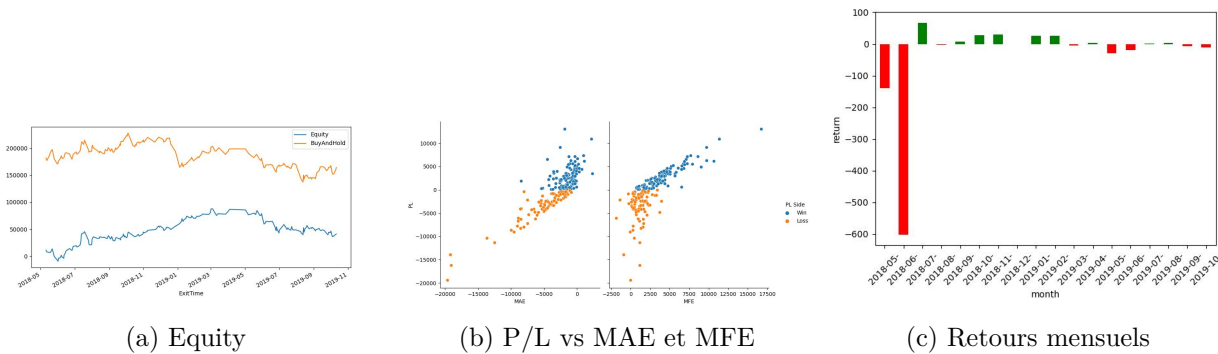


FIGURE C.9 – Résultats pour USD/JPY

Observation : Le modèle offre des résultats impressionnants avec un AVG Win/Loss exceptionnel (275.69) et un profit factor solide (1.19), malgré une accuracy inférieure à 50%. Cette robustesse du

modèle pourrait être une illusion due aux importants taux de changes entre le Yen et l'US dollar, et nécessite donc une analyse approfondie.