

---

# STRATÉGIE DE TRADING UTILISANT LE MACHINE LEARNING SUR L'ACTION AMAZON

---

*par*

**Khadija Masmoudi**

**Soufian Barkati**

**Serigne Saliou Mbacke**

**Ghita Khaloui**

Sous la direction de

**M. Dustin Lamblin**

2023-2024

Tous droits réservés©

## Table des matières

I	Introduction	2
II	Données et Caractéristiques	3
III	Méthodologie	4
IV	Modèles de Machine Learning	4
V	Choix des Métriques	4
VI	Résultats	5
VII	Discussion	5
VIII	Optuna	6
IX	Benchmark	7
X	Back-testing	7
XI	Conclusion	8

# **I Introduction**

---

Le trading algorithmique, en particulier lorsqu'il est renforcé par des techniques de Machine Learning, présente une opportunité sans précédent de comprendre et d'exploiter les dynamiques financières. Ce rapport explore l'élaboration d'une stratégie de trading exclusivement longue (long only) sur l'action Amazon (AMZN), en s'appuyant sur divers algorithmes de classification.

À travers l'analyse des données historiques d'Amazon, nous cherchons à déchiffrer le comportement de cette stratégie dans un contexte réel et à évaluer son efficacité. En intégrant des transformations multiples et des caractéristiques dérivées du prix de clôture, nous visons à établir une approche quantitative pour la prise de décision en trading.

## II Données et Caractéristiques

Ce projet utilise un ensemble de données sur les prix de l'action Amazon pour développer une stratégie de trading. Les données ont été prétraitées de la manière suivante :

- Conversion des dates au format standard et définition comme index.
- Application de différents indicateurs techniques comme SMA, EMA, RSI, et les bandes de Bollinger pour enrichir notre ensemble de données avec des caractéristiques susceptibles de capturer des tendances du marché.
- Normalisation des prix de clôture à l'aide de la moyenne mobile exponentielle et de l'écart-type mobile exponentiel pour calculer le score Z.

Des visualisations telles que des graphiques linéaires et des histogrammes ont été utilisées pour explorer la distribution des retours et la dynamique des prix. Le processus de sélection des caractéristiques a été guidé par l'importance de minimiser la redondance tout en capturant l'essence des tendances du marché pour la prise de décision de trading.

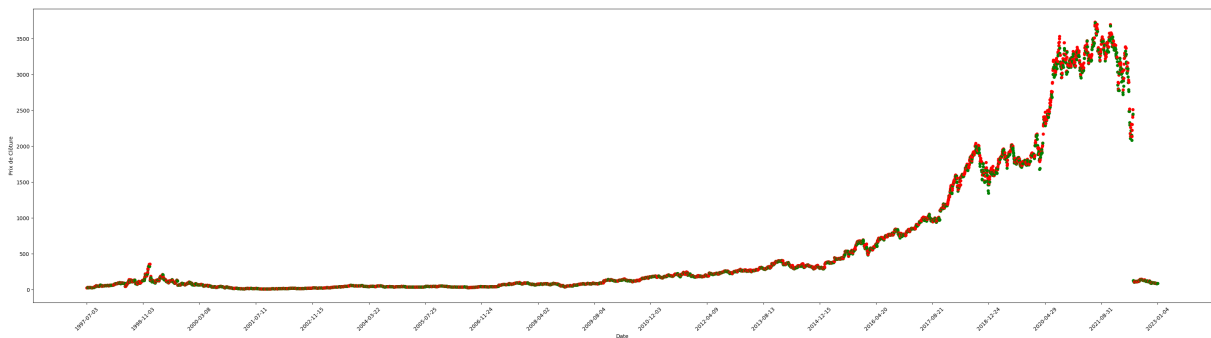


FIGURE 1 – Visualisation de la Dynamique du Marché : États Haussiers en Vert et Corrections en Rouge

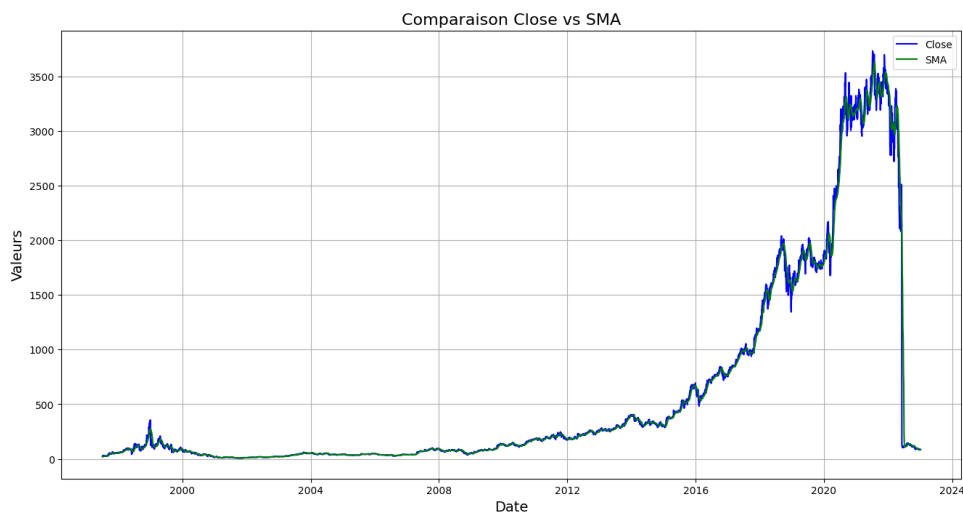


FIGURE 2 – Traçage prix de cloture et des indicateurs

### III Méthodologie

---

La méthodologie adoptée pour le prétraitement des données s’est appuyée sur plusieurs étapes critiques pour assurer l’efficacité de l’apprentissage automatique. Initialement, une conversion des dates au format `datetime` a été effectuée pour faciliter la manipulation temporelle des données. La définition de l’index comme la colonne de date a été une décision clé pour le traitement ultérieur.

Les principales méthodes de prétraitement incluaient la normalisation des données et l’ajout d’indicateurs techniques tels que **SMA**, **RSI**, et les **Bandes de Bollinger** pour enrichir notre ensemble de données avec des caractéristiques pertinentes. Le calcul des rendements quotidiens a également été effectué, suivi d’une transformation pour obtenir un label binaire indiquant une augmentation ou une diminution du prix de clôture.

Ces choix méthodologiques étaient motivés par la nécessité d’une représentation complète de la dynamique du marché et l’importance de capturer les tendances et les inversions potentielles. L’application de ces techniques de prétraitement était essentielle pour préparer les données pour la phase de modélisation et pour assurer que les modèles de machine learning puissent interpréter efficacement les motifs sous-jacents dans les mouvements de prix d’Amazon.

### IV Modèles de Machine Learning

---

Dans cette section, nous explorons les différents modèles de Machine Learning utilisés pour la stratégie de trading sur l’action Amazon. Nous avons opté pour des modèles variés tels que la régression logistique, l’arbre de décision et la forêt aléatoire. Chaque modèle a été sélectionné pour sa pertinence par rapport au type de données et aux objectifs de prédiction. Les paramètres de chaque modèle ont été finement ajustés pour optimiser les performances, en s’appuyant sur des techniques de validation croisée et de recherche de grille pour la sélection des hyperparamètres.

### V Choix des Métriques

---

Nous avons utilisé des méthodes telles que la précision, le rappel et le score F1 pour évaluer nos modèles de classification. La précision est cruciale pour minimiser les faux positifs, c’est-à-dire les occasions où le modèle prédit à tort un mouvement de prix favorable. Notre modèle permet de capturer une grande partie des mouvements réels positifs grâce au rappel, ou sensibilité. Le score F1 fournit un équilibre entre la précision et le rappel, ce qui est crucial pour gérer les situations de déséquilibre de classe où les mouvements de prix significatifs sont rares par rapport aux périodes de stabilité.

La matrice de confusion offre une vue d'ensemble des prédictions correctes et incorrectes du modèle, distinguant les vrais positifs, vrais négatifs, faux positifs et faux négatifs. Cette répartition nous permet de mesurer précisément la précision, le rappel, et d'autres métriques spécifiques au domaine du trading.

En traçant le taux de vrais positifs contre le taux de faux positifs à différents seuils de classification, la courbe ROC offre une analyse complète de la performance du modèle à divers niveaux de sensibilité. Une aire sous la courbe ROC (AUC) élevée indique que le modèle a une bonne capacité à différencier les mouvements de prix positifs des négatifs, ce qui est crucial pour optimiser notre stratégie de trading.

## VI Résultats

---

Les résultats obtenus après l'application des modèles sélectionnés sont présentés en détail dans cette section. Nous évaluons la performance de chaque modèle en utilisant des métriques clés telles que la précision, le rappel et l'aire sous la courbe ROC. Les analyses des erreurs et des caractéristiques importantes sont également discutées pour comprendre les forces et les faiblesses de notre approche de trading algorithmique.

Les résultats des modèles sont les suivants :

Modèle	Score d'Entraînement	Score de Validation	Score de Test
Arbre de Décision	0.613	0.605	0.604
Régression Logistique	0.597	0.623	0.604
Forêt Aléatoire	0.921	0.788	0.897

La forêt aléatoire affiche le score le plus élevé sur l'ensemble d'entraînement.

Sur la base de ces résultats, le Random Forest semble être un meilleur choix en termes de capacité à généraliser sur de nouvelles données.

## VII Discussion

---

Cette partie permet de mieux comprendre les performances des modèles de machine learning et d'identifier les forces, les limitations et les pistes d'amélioration pour des futures recherches.

Tout d'abord, il est important de noter que les modèles ont montré des performances variables selon les métriques utilisées. Le Random Forest a affiché des scores de précision et de rappel élevés sur l'ensemble d'entraînement, mais a eu des performances variables sur l'ensemble de test, indiquant une certaine difficulté à généraliser les prédictions à de nouvelles données.

Une des principales limitations de cette étude est la taille limitée de l'échantillon de données. Bien que nous ayons pu obtenir des résultats prometteurs, il est possible que

les performances des modèles puissent être améliorées avec un ensemble de données plus vaste et plus varié. De plus, le choix des caractéristiques et des hyperparamètres des modèles peut également avoir un impact significatif sur les performances, et des recherches supplémentaires pourraient être menées pour explorer différentes combinaisons de caractéristiques et de paramètres.

En outre, il est important de noter que les performances des modèles peuvent varier en fonction de l'évolution des conditions du marché. Des recherches futures pourraient inclure une analyse plus approfondie des facteurs macroéconomiques et des événements spécifiques au secteur qui pourraient influencer les prédictions des modèles.

## VIII Optuna

Optuna a permis de recueillir des informations précieuses sur le choix du modèle et l'ajustement des hyperparamètres pour notre stratégie de trading algorithmique. Les graphiques d'Optuna montrent que la forêt aléatoire dépasse les autres classificateurs, avec un rappel optimisé comme critère clé. Les variations de profondeur maximale et de poids des classes dans les modèles de forêt aléatoire montrent que des réglages plus fins peuvent améliorer significativement les performances.

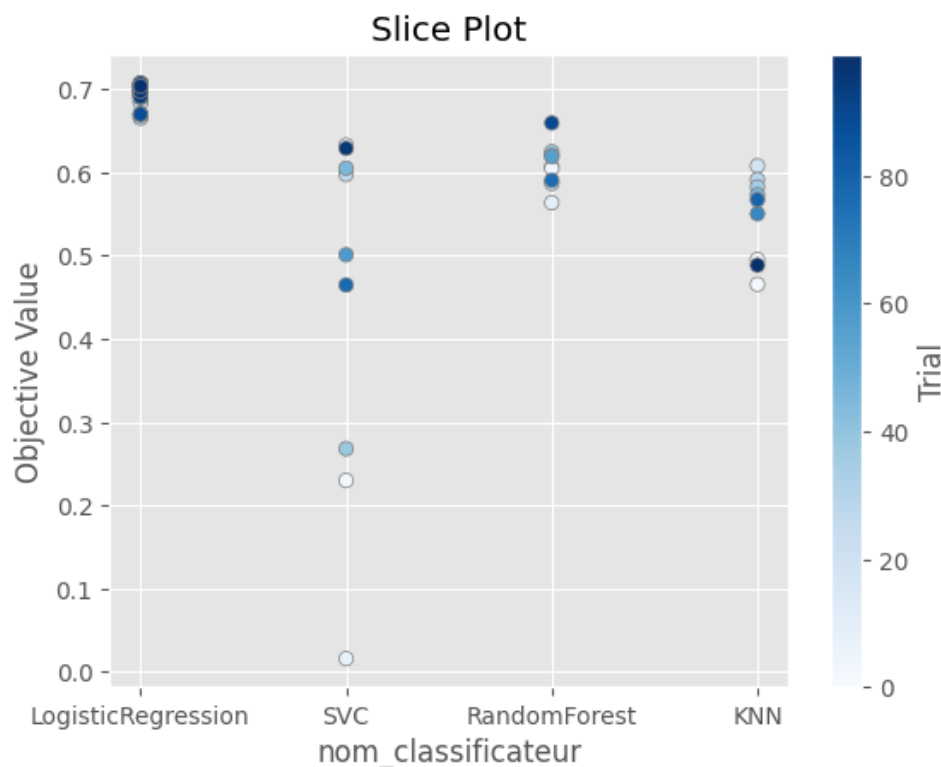


FIGURE 3 – Graphique 2D de la fonction objectif par rapport aux hyperparamètres pour le paramètre "nom\_classificateur"

## IX Benchmark

Nous avons établi un benchmark pour notre modèle de forêt aléatoire en utilisant un `DummyClassifier` avec une stratégie uniforme comme modèle de référence. Le `DummyClassifier` attribue des labels de façon aléatoire, reflétant ainsi une performance de base sans discernement. Les métriques de précision, rappel et score F1, tous avoisinant 0,49, nous montrent que cette performance aléatoire est équivalente à une chance de 50/50 dans la classification des données de test. Cela met en perspective la supériorité de notre modèle, qui doit dépasser ce seuil de performance pour être considéré comme ayant appris des motifs pertinents des données.

## X Back-testing

Notre stratégie de trading a été évaluée à travers un processus de back-testing approfondi, comparant les rendements cumulés de notre approche algorithmique à une stratégie d'achat et de conservation. Certaines tendances bénéfiques ont été capturées par notre stratégie, mais nous avons également observé des périodes de volatilité qui suggèrent une gestion plus fine du risque et un ajustement des paramètres du modèle. En somme, ces résultats soulignent l'importance d'une analyse continue pour affiner la stratégie et souligner le potentiel d'amélioration en alignant plus étroitement le modèle sur les dynamiques du marché.

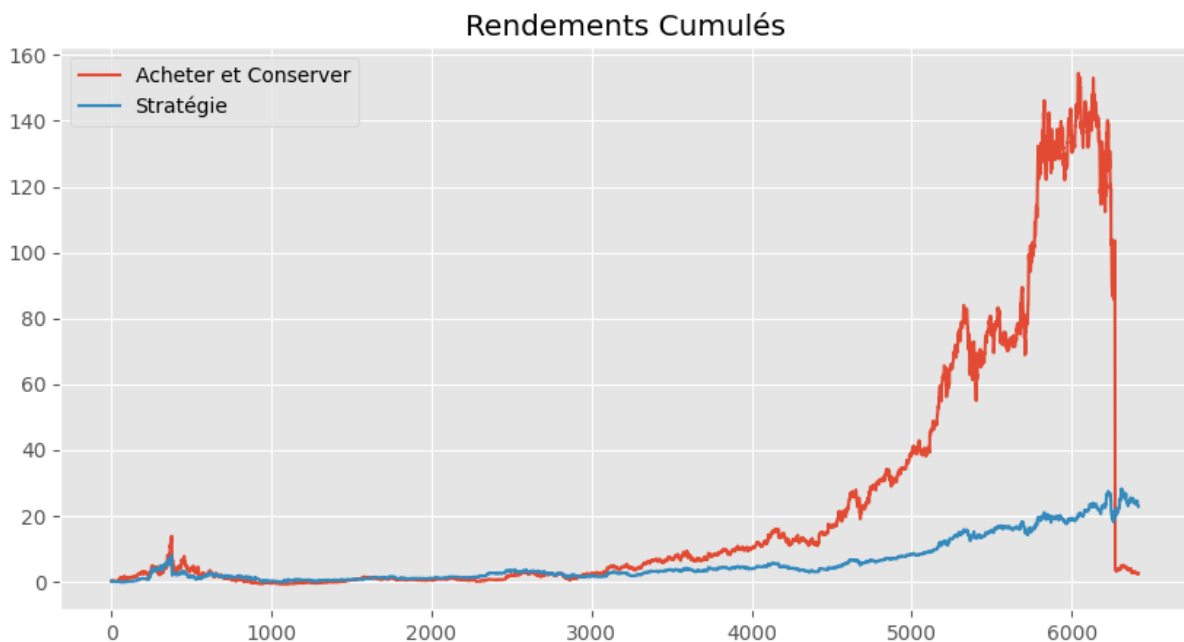


FIGURE 4 – Rendements Cumulés



## XI Conclusion

---

Cette étude a exploré l'application de modèles de machine learning à la stratégie de trading sur l'action Amazon. Les résultats obtenus ont permis de tirer plusieurs conclusions importantes.

Tout d'abord, nous avons constaté que différents modèles de machine learning présentent des performances variables en termes de prédiction des tendances du marché boursier. La régression logistique et la forêt aléatoire ont montré des performances prometteuses sur les ensembles d'entraînement, mais ont eu des difficultés à généraliser correctement les prédictions à de nouvelles données. L'arbre de décision, bien que moins performant en termes de précision et de rappel, a montré une stabilité relative dans ses performances sur les ensembles d'entraînement, de validation et de test.

De plus, nous avons identifié plusieurs limitations et pistes d'amélioration pour de futures recherches. La taille limitée de l'échantillon de données et le choix des caractéristiques et des hyperparamètres des modèles sont parmi les principales limitations à prendre en compte. Des recherches futures pourraient également explorer l'impact des conditions macroéconomiques et des événements spécifiques au secteur sur les performances des modèles.

Enfin, nous recommandons une approche prudente dans l'utilisation de modèles de machine learning pour la prise de décision en matière de trading. Bien que ces modèles puissent fournir des indications utiles sur les tendances du marché, il est important de les considérer comme des outils complémentaires à l'analyse fondamentale et à l'expertise humaine. Une validation régulière des performances des modèles et une adaptation aux changements du marché sont essentielles pour garantir le succès d'une stratégie de trading basée sur l'apprentissage automatique.