

# Prédiction des variations du cours du Bitcoin : Une approche basée sur l'analyse de données temporelles

Didier Henry

LAMIA, Université des Antilles, Guadeloupe, France  
didier.henry@univ-antilles.fr

**Résumé.** Dans cet article, nous explorons la complexité des cryptomonnaies en nous penchant sur la prédiction du cours du Bitcoin, un enjeu capital pour les investisseurs. En utilisant une analyse des corrélations entre le prix du Bitcoin et divers indicateurs, nous avons élaboré quatre modèles de prédiction. Le modèle forêt aléatoire s'est révélé pertinent, car il obtient une prédiction du cours du Bitcoin à 18 dollars près. Dans le cadre du défi proposé par la conférence EGC 2024, cette étude a également révélé des corrélations inattendues entre les anomalies entre les transactions de différents acteurs et les variations du prix du Bitcoin. L'analyse du réseau de transactions a également mis en lumière des tendances surprenantes, telles que l'augmentation du coefficient de clustering en période de croissance du Bitcoin et la diminution du nombre de communautés lors de ces hausses.

## 1 Introduction

Le Bitcoin est une cryptomonnaie révolutionnaire qui a suscité un intérêt croissant depuis sa création (Nakamoto, 2008). Son cours est caractérisé par une forte volatilité (Yermack, 2015). Dans la littérature, de nombreuses études ont été menées (Bouoiyour et Selmi, 2015; Zohar, 2015; Antonopoulos, 2014) pour comprendre les mécanismes sous-jacents du cours du Bitcoin et pour prédire ses variations. D'autres travaux (Velankar et al., 2018; Khedr et al., 2021) ont utilisé une variété de modèles et de données, offrant des perspectives intéressantes pour décrire le comportement de cette cryptomonnaie. Dans le cadre du défi proposé par la conférence "*Extraction et Gestion des Connaissances*" (EGC) 2024, nous avons exploré la prédiction du cours du Bitcoin en utilisant une approche multidimensionnelle. Notre travail s'appuie sur une analyse de données réelles variées, comprenant des données relatives au cours du Bitcoin entre le 01/01/2015 et le 30/06/2017. Les résultats que nous présentons dans cet article offrent un aperçu des tendances que nous avons identifiées, des corrélations significatives entre les attributs, et des modèles de prédiction que nous avons utilisés. L'article est structuré de la façon suivante. La section 2 décrit en détail les corrélations entre le cours du Bitcoin et divers attributs, en utilisant deux ensembles de données distincts. La Section 3 est consacrée à la détection d'anomalies dans les données de transactions entre des acteurs de la blockchain. La section 4 est, elle, dédiée à l'analyse du graphe de transactions de la blockchain. Enfin, la section 5 conclut notre travail et décrit les pistes de futures recherches dans ce domaine dynamique.

Prédiction des variations du cours du Bitcoin : Une approche basée sur l'analyse de données temporelles

## 2 Prédiction du cours du Bitcoin

Le problème de prédiction du Bitcoin est un problème plutôt complexe. En effet, le Bitcoin peut être influencé par de nombreux facteurs extérieurs (inflation, crises financières, politiques économiques des gouvernements, etc). Cependant, il peut être formalisé de la façon suivante :  $Y_{t+1} = f(X_1, X_2, \dots, X_n, t)$  où :

- $Y_{t+1}$  représente le cours du Bitcoin du jour suivant
- $X_1, X_2, \dots, X_n$  sont les  $N$  variables explicatives mesurées à l'instant  $t$ .

Dans le cadre du défi proposé par la conférence EGC, plusieurs jeux de données réelles ont été fournis. Dans notre approche, nous avons utilisé les fichiers *externals.csv* et *blockchain-global.csv* qui contiennent respectivement des données relatives au cours du Bitcoin (PriceUSD) et des données agrégées et calculées à partir des données de la blockchain Bitcoin. De plus, nous avons généré un fichier *externals\_with\_indicators.csv* dans lequel nous avons calculé de nouveaux attributs souvent utilisés en économie à partir du fichier *externals.csv* :

- *ROI* : Le rendement sur investissement.
- *PriceUSD\_MA7* : La moyenne mobile sur 7 jours du cours du Bitcoin.
- *PriceUSD\_MA30* : La moyenne mobile sur 30 jours du cours du Bitcoin.
- *Volatility* : La volatilité du cours du Bitcoin.
- *PriceUSD\_Diff\_MA7* : La différence entre le cours actuel et la moyenne mobile sur 7 jours.
- *MACD* : La convergence et divergence de la moyenne mobile.
- *RSI* : L'indice de force relative.

Dans un premier temps, nous avons observé les corrélations entre la valeur de PriceUSD et les nouveaux attributs que nous avons calculés. Nous avons sélectionné les attributs les plus corrélés avec PriceUSD pour l'entraînement des modèles prédictifs à savoir : "*HashRate\_J-1*", "*PriceUSD\_MA7\_J-1*", "*PriceUSD\_MA30\_J-1*", "*Volatility\_J-1*" et "*MACD\_J-1*". Les valeurs des corrélations des attributs du jour  $J - 1$  et  $J - 2$  étant proches, nous avons choisi de ne retenir que les attributs du jour  $J - 1$  dans la tâche de prédiction de l'attribut PriceUSD. En utilisant ces attributs, nous avons entraîné 4 modèles dans l'objectif de prédire le cours du Bitcoin : un modèle de régression linéaire (Linear Regression), un modèle de forêt aléatoire (*Random Forest*), un modèle de *Boosting Gradient* et un modèle de réseaux de neurones profonds (*DNN*). Nous avons noté que c'est un modèle de forêt aléatoire qui donne les meilleurs résultats (voir Figure 1).

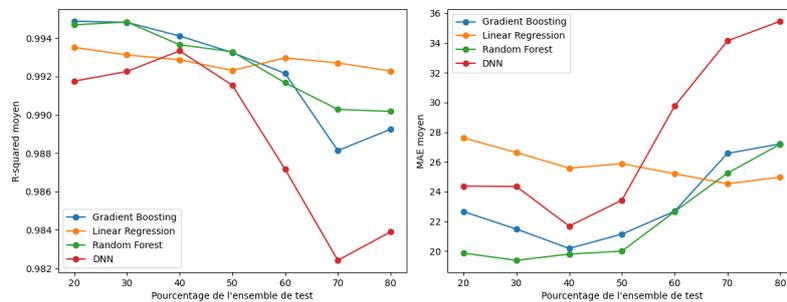


FIG. 1 – Scores des différents modèles en fonction de la taille de l'ensemble de test

Pour évaluer la robustesse des modèles, nous avons effectué une validation croisée en divisant l'ensemble des données en 5, une valeur choisie arbitrairement. En observant la figure 2, nous constatons que c'est le modèle forêt aléatoire qui obtient les meilleurs résultats avec une moyenne des erreurs absolues égale à 18 \$.

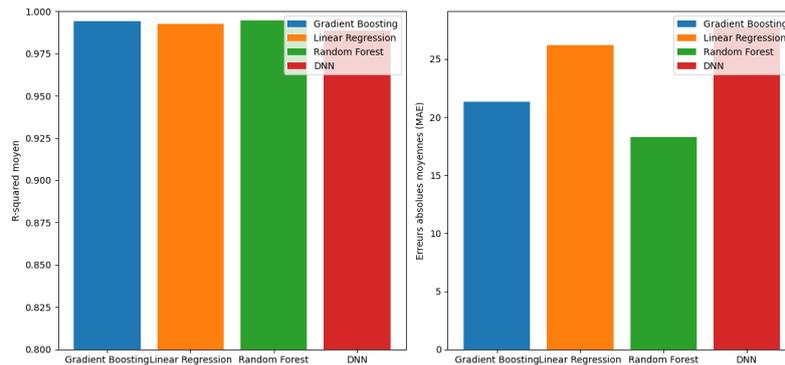


FIG. 2 – Scores des différents modèles en fonction de la taille de l'ensemble de test

Dans un second, nous nous sommes intéressés aux corrélations entre la valeur de PriceUSD et les attributs du fichier *blockchain\_global.csv*. En suivant la même méthode que précédemment, nous avons sélectionné les plus corrélés avec PriceUSD. Ainsi les attributs suivants ont été choisis : "Total\_Fee\_J-1", "mean\_fee\_satoshi\_J-1", "mean\_feeUSD\_J-1" et "nb\_transactions\_J-1". Comme précédemment, nous avons utilisé les mêmes 4 modèles pour la prédiction du cours du Bitcoin. Nous avons noté que c'est encore le modèle forêt aléatoire qui donne les meilleurs résultats. Cependant, nous avons également constaté que les performances sont généralement moins bonnes en utilisant ces attributs (voir figure 3).

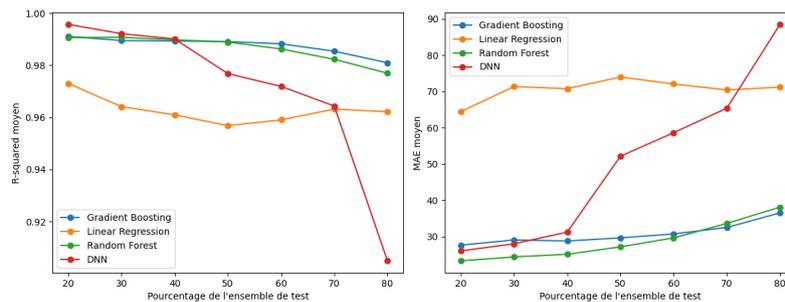


FIG. 3 – Scores des différents modèles en fonction de la taille de l'ensemble de test

Ensuite, nous avons effectué une validation croisée de façon analogue à la première. D'une part, nous avons remarqué une baisse des performances des modèles. D'autre part, nous avons constaté que le modèle de réseaux de neurones donne est le moins adapté en utilisant ces attributs (voir figure 4).

## Prédiction des variations du cours du Bitcoin : Une approche basée sur l'analyse de données temporelles

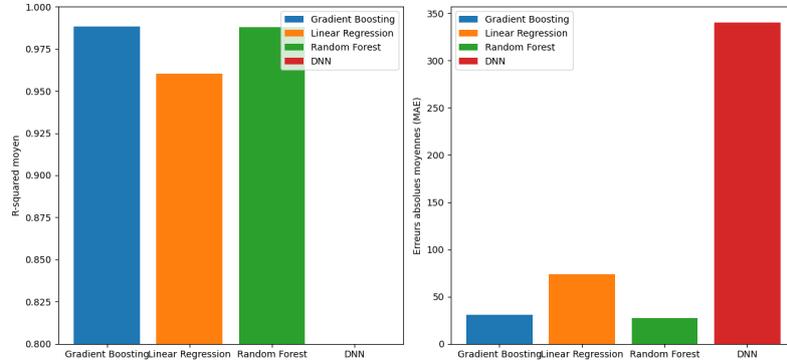


FIG. 4 – Scores des différents modèles en fonction de la taille de l'ensemble de test

Enfin, nous avons regroupé l'ensemble des attributs sélectionnés à la première analyse et à la deuxième analyse puis nous avons entraîné les 4 modèles. En observant les résultats obtenus, nous n'avons pas noté d'amélioration significative par rapport aux performances obtenues lors de la première analyse. Néanmoins, nous observons que c'est à chaque fois le modèle forêt aléatoire qui donne les meilleurs résultats.

### 3 Détection d'anomalies

L'analyse des anomalies au niveau des transactions du Bitcoin pourrait révéler des informations cruciales sur les comportements inhabituels des acteurs du marché. Ces anomalies pourraient être des signes précurseurs de mouvements significatifs du cours du Bitcoin. Par conséquent, en identifiant et en comprenant ces anomalies, les investisseurs et les acteurs du marché pourraient prendre des mesures préventives pour minimiser les risques ou même exploiter les opportunités qui se présentent. Dans cette partie, nous allons présenter en détail la méthode de détection d'anomalies que nous avons appliquée au fichier "*blockchain\_by\_actor.csv*" et analyser les résultats obtenus en mettant en lumière les correspondances entre les anomalies et les variations du cours du Bitcoin. Premièrement nous avons cherché à identifier des événements qui se démarquent significativement des comportements habituels des acteurs des transactions du Bitcoin. Pour ce faire, nous avons utilisé une technique basée sur la déviation par rapport à la moyenne mobile. Cela peut être formalisé de la manière suivante. Soit  $transactions(t)$  la série temporelle du nombre de transactions effectuées par un acteur donné, où  $t$  représente le temps. Nous avons calculé la moyenne mobile sur 7 jours de  $transactions(t)$  comme suit :

$$rolling\_mean(t) = \frac{1}{7} \sum_{i=1}^7 transactions(t - i)$$

Nous avons calculé l'écart-type de cette moyenne mobile pour quantifier la variabilité :

$$std\_dev = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (transactions(t) - rolling\_mean(t))^2}$$

Nous avons défini un seuil en fonction de l'écart-type :  $threshold = 2 * std\_dev$ . Enfin nous avons identifié les dates d'anomalies lorsque la différence absolue entre le nombre de transactions réel et la moyenne mobile dépasse ce seuil :

$$anomalies(t) = \{\|transactions(t) - rolling\_mean(t)\| > threshold\}$$

Par exemple, dans la figure 5, nous pouvons voir l'application de cette méthode de détection d'anomalies dans le nombre de transactions pour un des acteurs présent dans le jeu de données.

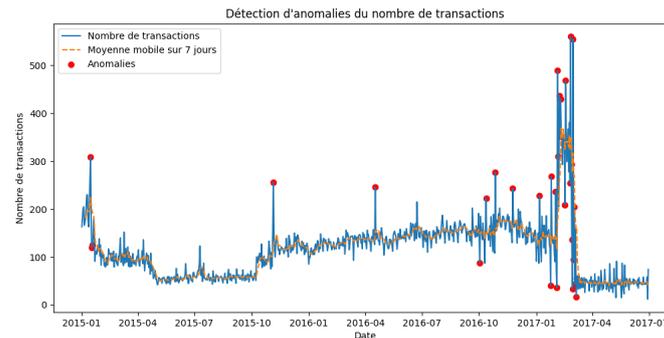


FIG. 5 – Détection d'anomalies du nombre de transactions pour l'acteur *BTC-e.com*

Pour compléter notre analyse, nous avons identifié l'ensemble des dates des anomalies pour chacun des acteurs et nous avons sélectionné les 20 dates les plus fréquentes parmi l'ensemble de ces dates. Ainsi, nous avons observé que les anomalies du nombre de transactions détectées semblent correspondre aux fortes augmentations du cours du Bitcoin (voir Figure 6). Par exemple, les anomalies détectées fin 2016 et début 2017 peuvent être liées à la chute de la valeur de la monnaie chinoise, l'élection de D.Trump et au Brexit.

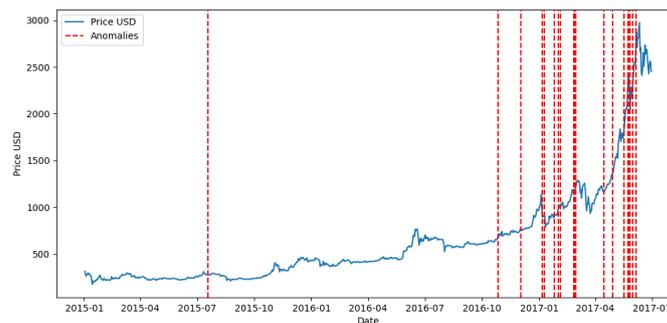


FIG. 6 – Evolution du cours du Bitcoin avec les 20 anomalies les plus fréquentes détectées du nombre de transactions

De la même façon, nous avons observé que les anomalies concernant les attributs *sum\_fee* et *mean\_fee\_for\_100* semblent également correspondre aux fortes augmentations du cours du Bitcoin. D'autre part, nous nous sommes intéressés aux anomalies de l'attribut "*self\_spent*"

Prédiction des variations du cours du Bitcoin : Une approche basée sur l'analyse de données temporelles

qui correspond aux montants observés comme envoyés de l'acteur à lui-même. Cela correspondrait à un comportement atypique ou assez suspect de l'acteur, car il envoie soudainement une quantité anormale de fonds à lui-même. Dans la figure 7, nous avons identifié le top 10 des acteurs ayant le plus d'anomalies.

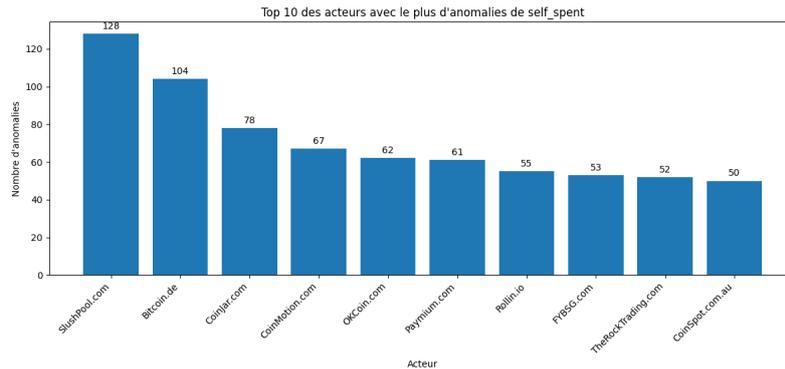


FIG. 7 – Top 10 des acteurs avec le plus d'anomalies de "self\_spent"

## 4 Analyses du graphe de transactions

Dans cette partie, nous nous intéressons à l'analyse du graphe des transactions au cours du temps. Alors que le Bitcoin évolue en tant que système financier décentralisé, son réseau de transactions devient un terrain fertile pour une compréhension approfondie et une exploration des tendances émergentes. Il semble de plus en plus clair que le comportement du réseau de transactions du Bitcoin est étroitement lié à son cours et à ses variations. À mesure que le marché des cryptomonnaies continue de mûrir, il est essentiel de plonger dans l'examen des caractéristiques dynamiques de ce réseau. Cette analyse peut offrir des aperçus précieux sur les mécanismes sous-jacents du Bitcoin. Nous nous sommes intéressés à différentes métriques du graphe des transactions, telles que le nombre de noeuds, le nombre d'arêtes, le nombre de nouveaux noeuds, le nombre de nouvelles arêtes, le coefficient de clustering, la densité du réseau, ou encore le nombre de communautés. Dans notre approche, nous avons choisi le coefficient de clustering dans sa version locale, évaluant ainsi la connectivité des voisins directs d'un noeud dans le réseau. Pour identifier les structures communautaires, nous avons utilisé la bibliothèque *NetworkX* en Python, reconnue pour sa capacité à analyser les propriétés structurelles des réseaux. Premièrement, nous avons observé que le coefficient de clustering augmente quand le cours du Bitcoin augmente (voir Figure 8). En fait, une augmentation du coefficient de clustering pourrait indiquer une adoption accrue du Bitcoin. Lorsque le prix du Bitcoin augmente, de plus en plus de personnes et d'entités peuvent être incitées à rejoindre le réseau, entraînant ainsi une augmentation du nombre de connexions entre les utilisateurs. Cela crée un réseau de transactions plus dense, ce qui se reflète dans le coefficient de clustering plus élevé.

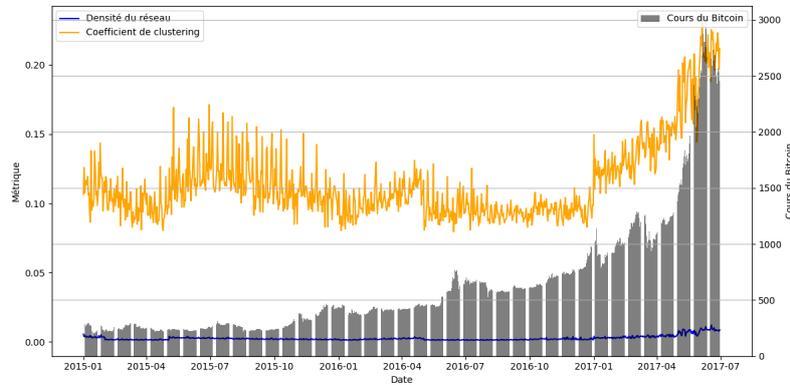


FIG. 8 – Evolution du coefficient de clustering et de la densité du réseau en fonction du temps.

Dans un second temps, nous avons constaté que le nombre de communautés semble diminué quand le cours de Bitcoin augmente (voir Figure 9). Cela pourrait signifier que si le Bitcoin gagne en valeur, les effets de réseau deviennent plus forts. Les acteurs peuvent être plus enclins à se conformer aux normes et aux pratiques établies par la majorité, ce qui peut contribuer à la réduction du nombre de communautés.

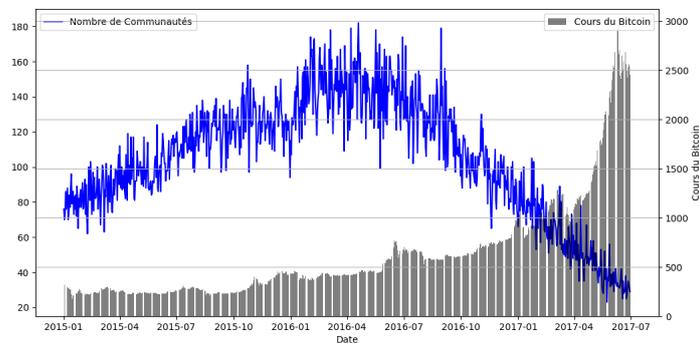


FIG. 9 – Evolution du nombre de communautés du réseau en fonction du temps.

## 5 Conclusion

En conclusion dans ce travail, nous avons mis l'accent sur la prédiction du cours du Bitcoin, une tâche cruciale pour les investisseurs. À travers une analyse approfondie des corrélations entre le prix du Bitcoin et divers indicateurs, nous avons élaboré quatre modèles de prédiction. Parmi eux, le modèle forêt aléatoire a montré des prédictions du cours du Bitcoin avec une précision proche de 18 dollars. Dans le cadre du défi de la conférence EGC 2024, notre étude a révélé des liens inattendus entre les anomalies dans les transactions des acteurs et les variations

Prédiction des variations du cours du Bitcoin : Une approche basée sur l'analyse de données temporelles

du prix du Bitcoin. De plus, notre exploration du réseau de transactions a mis en lumière des tendances surprenantes, telles que l'augmentation du coefficient de clustering lors des périodes de croissance du Bitcoin et la diminution du nombre de communautés lors de ses hausses. Bien que ces résultats soient intéressants, ils soulignent également la persistance de la complexité du cours du Bitcoin. Cependant, ils ouvrent la voie à une compréhension plus approfondie et à des investissements plus éclairés, représentant un avenir prometteur pour l'analyse des données temporelles dans la prédiction des marchés financiers.

## Références

- Antonopoulos, A. M. (2014). *Mastering Bitcoin : unlocking digital cryptocurrencies*. " O'Reilly Media, Inc."
- Bouoiyour, J. et R. Selmi (2015). What does bitcoin look like? *Annals of Economics & Finance* 16(2).
- Khedr, A. M., I. Arif, M. El-Bannany, S. M. Alhashmi, et M. Sreedharan (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques : A survey. *Intelligent Systems in Accounting, Finance and Management* 28(1), 3–34.
- Nakamoto, S. (2008). Bitcoin : A peer-to-peer electronic cash system. *Decentralized business review*.
- Velankar, S., S. Valecha, et S. Maji (2018). Bitcoin price prediction using machine learning. In *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pp. 144–147. IEEE.
- Yermack, D. (2015). Is bitcoin a real currency? an economic appraisal. In *Handbook of digital currency*, pp. 31–43. Elsevier.
- Zohar, A. (2015). Bitcoin : under the hood. *Communications of the ACM* 58(9), 104–113.

## Summary

In this article, we explore the complexity of cryptocurrencies by focusing on predicting the price of Bitcoin, a key issue for investors. Using correlation analysis between Bitcoin price and various indicators, we developed four prediction models. The Random Forest model proved to be relevant, as it obtained a prediction of the Bitcoin price to within 18 dollars. As part of the challenge proposed by the EGC 2024 conference, this study also revealed unexpected correlations between anomalies between transactions of different actors and variations in the price of Bitcoin. Analysis of the transaction network also revealed surprising trends, such as the increase in the clustering coefficient during periods of Bitcoin growth and the decrease in the number of communities during its rises.