

# **Analyse des Cryptomonnaies et Modélisation Prédictive du Bitcoin**

Université Hassan 1er – ENCG Settat

Rapport encadré par Mr. LARHLIMI

Étudiant : Imran Elboti

December 4, 2025

# Contents

<b>1</b>	<b>Introduction Générale aux Cryptomonnaies</b>	<b>3</b>
1.1	Contexte Historique et Évolution . . . . .	3
1.2	Fondements Technologiques : La Blockchain . . . . .	3
1.3	Dynamiques Socio-économiques d'Adoption . . . . .	4
1.4	Importance de l'Analyse Quantitative . . . . .	4
<b>2</b>	<b>Focus sur le Bitcoin</b>	<b>5</b>
2.1	Position Dominante dans l'Écosystème . . . . .	5
2.2	Architecture Technique et Mécanisme de Consensus . . . . .	5
2.3	Dynamiques de Volatilité et Facteurs d'Influence . . . . .	6
2.4	Données Disponibles et Potentiel d'Analyse . . . . .	6
<b>3</b>	<b>Présentation de la Base de Données</b>	<b>7</b>
3.1	Source et Caractéristiques Générales . . . . .	7
3.2	Variables Disponibles et Définitions . . . . .	7
3.3	Justification du Focus sur le Bitcoin . . . . .	7
3.4	Pré-analyse et Observations Préliminaires . . . . .	8
<b>4</b>	<b>Méthodologie de Préparation et de Transformation des Données</b>	<b>9</b>
4.1	Processus de Nettoyage et de Validation . . . . .	9
4.2	Gestion des Données Manquantes et Aberrantes . . . . .	9
4.3	Standardisation et Normalisation . . . . .	10
4.4	Création de Variables Déduites et Techniques Feature Engineering . . . . .	10
4.5	Séparation des Données et Validation Temporelle . . . . .	10

<b>5 Méthode de Modélisation : Régression et Apprentissage Automatique</b>	<b>12</b>
5.1 Architecture Globale de la Modélisation . . . . .	12
5.2 Algorithmes Implémentés . . . . .	12
5.2.1 Régression Linéaire . . . . .	12
5.2.2 Random Forest Regressor . . . . .	13
5.2.3 Gradient Boosting Regressor . . . . .	13
5.3 Procédures de Validation et d'Optimisation . . . . .	13
5.4 Métriques d'Évaluation . . . . .	14
<b>6 Résultats et Analyse</b>	<b>15</b>
6.1 Comparaison des Performances Globales . . . . .	15
6.2 Analyse des Résidus . . . . .	15
6.2.1 Distribution des Résidus . . . . .	15
6.2.2 Graphiques des Résidus . . . . .	16
6.3 Importance des Variables . . . . .	17
6.4 Comparaison des RMSE par Période . . . . .	18
<b>7 Limites de la Méthode</b>	<b>19</b>
7.1 Limitations Structurelles des Modèles . . . . .	19
7.2 Problèmes Spécifiques aux Séries Temporelles Financières . . . . .	19
7.3 Limitations des Données Disponibles . . . . .	20
7.4 Problèmes de Généralisation Temporelle . . . . .	20
<b>8 Conclusion et Perspectives</b>	<b>21</b>
8.1 Synthèse des Principaux Résultats . . . . .	21
8.2 Contributions Méthodologiques . . . . .	21
8.3 Limites et Pistes d'Amélioration . . . . .	22
8.4 Implications Pratiques et Théoriques . . . . .	22

# Chapter 1

## Introduction Générale aux Cryptomonnaies

### 1.1 Contexte Historique et Évolution

Les cryptomonnaies représentent une innovation majeure dans l'histoire financière contemporaine. Apparues en 2009 avec le lancement du Bitcoin par le pseudonyme *Satoshi Nakamoto*, elles ont bouleversé les paradigmes traditionnels liés au contrôle monétaire, aux systèmes financiers centralisés et à la gouvernance économique. Cette innovation s'inscrit dans un contexte post-crise financière de 2008, marqué par une méfiance croissante envers les institutions financières traditionnelles.

L'émergence des cryptomonnaies n'est pas un phénomène isolé, mais s'intègre dans une transformation plus large des systèmes économiques et technologiques. Elle combine des avancées en cryptographie, en théorie des jeux et en réseaux distribués, créant ainsi une nouvelle classe d'actifs numériques décentralisés.

### 1.2 Fondements Technologiques : La Blockchain

Contrairement aux monnaies classiques, les cryptomonnaies reposent sur la technologie de la **blockchain**, un registre distribué, immuable et transparent permettant des transactions sécurisées sans l'intervention d'un tiers de confiance. La blockchain fonctionne comme un grand livre comptable public et décentralisé où chaque transaction est validée par consensus entre les noeuds du réseau.

Cette architecture distribuée présente plusieurs avantages fondamentaux : elle élimine le risque de point unique de défaillance, réduit les coûts de transaction, augmente la transparence des opérations et permet une résilience accrue face aux tentatives de censure

ou de manipulation.

### **1.3 Dynamiques Socio-économiques d’Adoption**

L’adoption des cryptomonnaies s’explique par trois dynamiques principales interconnectées. Premièrement, la recherche d’alternatives monétaires face aux systèmes financiers traditionnels perçus comme opaques et instables. Deuxièmement, l’évolution des technologies numériques qui rend possible la création et la gestion d’actifs décentralisés. Troisièmement, la croissance de l’économie numérique globale qui nécessite des moyens de paiement adaptés aux échanges transfrontaliers et aux nouvelles formes de valeur numérique.

Aujourd’hui, l’écosystème des cryptomonnaies dépasse plusieurs milliers d’actifs, chacun possédant ses caractéristiques, ses usages et sa structure économique. Cette diversité reflète l’expérimentation continue dans ce domaine, allant des monnaies de réserve de valeur comme le Bitcoin aux plateformes de contrats intelligents comme Ethereum, en passant par les tokens utilitaires et les stablecoins.

### **1.4 Importance de l’Analyse Quantitative**

Ce contexte riche et complexe justifie l’importance d’études statistiques et économétriques permettant de mieux comprendre leurs dynamiques, leur volatilité et leur potentiel prédictif. L’analyse quantitative des cryptomonnaies présente des défis uniques liés à leur haute volatilité, leur nouveauté relative et leur sensibilité aux facteurs tant techniques que réglementaires.

# Chapter 2

## Focus sur le Bitcoin

### 2.1 Position Dominante dans l'Écosystème

Le Bitcoin occupe une place centrale dans l'univers des cryptomonnaies. Il s'agit non seulement de la première cryptomonnaie créée, mais aussi de la plus utilisée et la plus capitalisée jusqu'à aujourd'hui. Sa position dominante en fait un indicateur de référence pour l'ensemble du marché des cryptomonnaies, influençant souvent les mouvements des autres actifs numériques.

L'objectif principal du Bitcoin est de proposer un système de paiement entièrement décentralisé, fondé sur le *peer-to-peer*, permettant des transferts de valeur sans intermédiaire entre n'importe quels participants du réseau, quels que soient leur localisation géographique ou leur statut juridique.

### 2.2 Architecture Technique et Mécanisme de Consensus

Sur le plan technique, le Bitcoin repose sur un algorithme de consensus appelé **Proof of Work** (PoW). Ce mécanisme nécessite une puissance de calcul importante pour valider les transactions et sécuriser le réseau. Les mineurs, participants qui mettent leur puissance de calcul à disposition du réseau, sont récompensés par de nouveaux bitcoins créés (subsidy) et par les frais de transaction.

L'offre de Bitcoin est limitée à 21 millions d'unités, une caractéristique déflationniste intentionnelle qui contraste avec les politiques monétaires expansionnistes de nombreuses banques centrales. Cette rareté programmée contribue à sa perception comme "or numérique" et réserve de valeur à long terme.

## 2.3 Dynamiques de Volatilité et Facteurs d’Influence

La volatilité du Bitcoin est un élément clé de son analyse. Contrairement aux actifs financiers traditionnels, ses fluctuations sont influencées par une combinaison unique de facteurs :

- **Facteurs réglementaires** : Les annonces gouvernementales et les décisions réglementaires internationales peuvent provoquer des mouvements brusques des prix.
- **Dynamiques de marché** : Les mouvements spéculatifs, les cycles d’adoption et les phénomènes de comportement de masse influencent les tendances à court et moyen terme.
- **Innovations technologiques** : Les mises à jour du protocole, les améliorations de scalabilité et les développements de l’écosystème secondaire impactent la perception de la valeur.
- **Adoption institutionnelle** : L’entrée des fonds d’investissement, des entreprises cotées et des institutions financières traditionnelles modifie la structure de la demande.

Cette volatilité exceptionnelle, bien que présentant des risques, en fait également un terrain d’application particulièrement intéressant pour les méthodes statistiques et de modélisation prédictive, permettant de tester la robustesse des modèles dans des conditions extrêmes.

## 2.4 Données Disponibles et Potentiel d’Analyse

La disponibilité de données historiques complètes et granulaires sur le Bitcoin, combinée à son importance systémique dans l’écosystème des cryptomonnaies, justifie son choix comme objet principal d’étude dans ce rapport. Son analyse peut servir de base pour comprendre des dynamiques applicables à d’autres actifs cryptographiques.

# Chapter 3

## Présentation de la Base de Données

### 3.1 Source et Caractéristiques Générales

La base de données analysée provient du dataset *Cryptocurrency Price History* disponible sur Kaggle. Ce dataset exhaustif contient l'historique complet des prix et volumes de plusieurs centaines de cryptomonnaies depuis leurs débuts. Pour ce rapport, nous avons choisi de concentrer l'analyse sur le Bitcoin afin d'obtenir une étude cohérente, précise et économiquement significative.

La période couverte par les données s'étend de la création du Bitcoin jusqu'à la date la plus récente disponible dans le dataset, offrant ainsi une perspective historique complète sur l'évolution de cet actif numérique.

### 3.2 Variables Disponibles et Définitions

La base de données contient les variables suivantes pour chaque date d'observation :

### 3.3 Justification du Focus sur le Bitcoin

Le choix de se concentrer exclusivement sur le Bitcoin dans cette analyse initiale se justifie par plusieurs raisons. Premièrement, le Bitcoin représente la cryptomonnaie la plus ancienne et la plus établie, offrant ainsi la série temporelle la plus longue et la plus complète. Deuxièmement, sa capitalisation dominante (généralement entre 40% et 60% du marché total) en fait un indicateur représentatif de l'ensemble du secteur. Troisièmement, la liquidité élevée du Bitcoin réduit les problèmes liés aux données clairsemées ou aux artefacts de prix.

### **3.4 Pré-analyse et Observations Préliminaires**

Une exploration initiale des données révèle les caractéristiques attendues du marché du Bitcoin : haute volatilité intra-journalière, croissance exponentielle à long terme interrompue par des corrections significatives, et corrélations intéressantes entre le volume des transactions et les mouvements de prix. Ces observations préliminaires guident le choix des méthodes de transformation et de modélisation.

# **Chapter 4**

## **Méthodologie de Préparation et de Transformation des Données**

### **4.1 Processus de Nettoyage et de Validation**

Avant toute modélisation, une préparation rigoureuse des données a été réalisée selon un processus structuré en plusieurs étapes. Cette phase préparatoire est cruciale pour garantir la qualité et la fiabilité des analyses ultérieures.

La première étape consiste en la conversion systématique des dates au format temporel approprié, permettant l'analyse chronologique et l'alignement temporel des différentes séries. Cette standardisation temporelle est essentielle pour les analyses de séries temporelles et la construction de variables décalées.

### **4.2 Gestion des Données Manquantes et Aberrantes**

La détection et suppression des doublons a été réalisée à l'aide d'algorithmes de comparaison exacte et approximative, garantissant l'unicité des observations. Les valeurs manquantes, bien que rares dans ce dataset, ont été traitées selon leur nature : les petites lacunes ont été comblées par interpolation linéaire, tandis que les périodes étendues sans données ont été exclues de l'analyse.

La validation de la cohérence interne des données a également été effectuée, vérifiant par exemple que les prix "High" sont toujours supérieurs ou égaux aux prix "Low", et que les prix de clôture se situent bien dans l'intervalle défini par ces extrêmes.

## 4.3 Standardisation et Normalisation

La standardisation des variables numériques a été réalisée grâce au **StandardScaler** de scikit-learn. Cette transformation centrée réduite, définie par :

$$X_{\text{standardisé}} = \frac{X - \mu}{\sigma}$$

où  $\mu$  est la moyenne et  $\sigma$  l'écart-type, permet de mettre toutes les variables sur la même échelle sans distordre leur distribution. Cette étape est particulièrement importante pour les algorithmes sensibles à l'échelle des variables comme la régression linéaire ou les méthodes à noyau.

## 4.4 Crédit de Variables Déduites et Techniques Feature Engineering

La création de nouvelles variables dérivées constitue une étape clé pour intégrer la dimension temporelle et capturer des dynamiques de marché non directement observables. Parmi les transformations appliquées :

- **Moyennes mobiles** : Calcul des moyennes mobiles sur 7, 14 et 30 jours pour lisser les fluctuations à court terme et identifier les tendances sous-jacentes.
- **Variables décalées** : Création de retards (lags) des prix de clôture sur différentes périodes pour modéliser l'autocorrélation temporelle.
- **Indicateurs techniques** : Calcul d'indicateurs standards comme les bandes de Bollinger, le RSI (Relative Strength Index) et les variations en pourcentage.
- **Caractéristiques temporelles** : Extraction des composantes de la date (jour de la semaine, mois, trimestre) pour capturer les effets saisonniers potentiels.

Ces transformations permettent aux modèles d'apprentissage automatique d'améliorer significativement leur capacité à capturer des tendances structurelles et des patterns récurrents dans les données.

## 4.5 Séparation des Données et Validation Temporelle

La séparation des données en ensembles d'entraînement et de test a été réalisée selon une approche chronologique stricte, préservant l'ordre temporel des observations. Cette

méthode évite le "data leakage" où des informations du futur pourraient contaminer les prédictions du passé.

Une validation croisée adaptée aux séries temporelles (**TimeSeriesSplit**) a été utilisée, avec des plis qui respectent la chronologie et évitent toute contamination temporelle entre les ensembles d'entraînement et de validation.

# Chapter 5

## Méthode de Modélisation : Régression et Apprentissage Automatique

### 5.1 Architecture Globale de la Modélisation

La méthode utilisée combine différents algorithmes de régression pour prédire le prix de clôture du Bitcoin selon une approche comparative systématique. L'objectif est non seulement d'obtenir des prédictions précises, mais aussi de comprendre les forces et faiblesses relatives de différentes familles d'algorithmes dans ce contexte spécifique.

Le pipeline de modélisation suit une architecture modulaire permettant l'évaluation indépendante de chaque algorithme tout en assurant la comparabilité des résultats grâce à des métriques standardisées et des procédures de validation identiques.

### 5.2 Algorithmes Implémentés

Trois modèles principaux ont été testés et comparés :

#### 5.2.1 Régression Linéaire

Le modèle de régression linéaire multivariée cherche à établir une relation linéaire entre les variables explicatives et le prix de clôture. Formellement, le modèle s'exprime comme :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

où  $y$  est le prix de clôture prédit, les  $x_i$  sont les variables explicatives, les  $\beta_i$  sont les coefficients à estimer, et  $\epsilon$  représente l'erreur.

Malgré sa simplicité, ce modèle fournit une base de référence utile et permet une interprétation directe des relations entre variables grâce aux coefficients estimés.

### 5.2.2 Random Forest Regressor

La méthode Random Forest (forêt aléatoire) est une technique d'ensemblage basée sur la construction de multiples arbres décisionnels. Chaque arbre est entraîné sur un échantillon bootstrap des données et un sous-ensemble aléatoire des variables. La prédiction finale est la moyenne des prédictions individuelles des arbres.

Cette approche présente plusieurs avantages : réduction du sur-apprentissage grâce au bagging, estimation naturelle de l'importance des variables, et capacité à capturer des relations non linéaires et des interactions complexes entre variables.

### 5.2.3 Gradient Boosting Regressor

Le Gradient Boosting est une technique d'ensemblage plus sophistiquée qui construit séquentiellement une série de modèles faibles (généralement des arbres de décision peu profonds), où chaque nouveau modèle corrige les erreurs des précédents. L'algorithme minimise une fonction de perte différentiable en utilisant la descente de gradient.

Particulièrement efficace pour les problèmes de régression, le Gradient Boosting excelle souvent dans les compétitions de data science grâce à sa capacité à capturer des patterns complexes tout en restant relativement robuste au sur-apprentissage avec un réglage approprié.

## 5.3 Procédures de Validation et d'Optimisation

Une validation croisée spécifique aux séries temporelles (**TimeSeriesSplit**) a été utilisée systématiquement pour éviter les problèmes liés à l'ordre chronologique des données. Cette méthode crée des plis où les données d'entraînement précèdent toujours chronologiquement les données de validation, simulant ainsi les conditions réelles de prédiction.

Ensuite, une recherche d'hyperparamètres via **GridSearchCV** a permis d'optimiser les performances de chaque modèle. Pour chaque algorithme, un espace de recherche défini systématiquement a été exploré :

- **Régression Linéaire** : Paramètres de régularisation (L1/L2), force de régularisation
- **Random Forest** : Nombre d'arbres, profondeur maximum, nombre minimum d'échantillons par feuille
- **Gradient Boosting** : Taux d'apprentissage, nombre d'estimateurs, profondeur maximum, fraction d'échantillons

## 5.4 Métriques d'Évaluation

Les performances des modèles ont été évaluées selon plusieurs métriques complémentaires :

- **RMSE (Root Mean Square Error)** : Mesure l'écart quadratique moyen entre prédictions et valeurs réelles, exprimée dans l'unité originale des prix.
- **MAE (Mean Absolute Error)** : Mesure l'écart absolu moyen, moins sensible aux valeurs extrêmes que le RMSE.
- **R<sup>2</sup> (Coefficient de détermination)** : Proportion de la variance expliquée par le modèle.
- **MAPE (Mean Absolute Percentage Error)** : Erreur en pourcentage, utile pour l'interprétation économique.

Ces métriques permettent une évaluation multidimensionnelle des performances, capturant différents aspects de la qualité prédictive.

# Chapter 6

## Résultats et Analyse

### 6.1 Comparaison des Performances Globales

L'évaluation comparative des trois modèles révèle des différences significatives dans leurs capacités prédictives. Le tableau ci-dessous présente les principales métriques d'évaluation :

Modèle	RMSE	MAE	R <sup>2</sup>	MAPE (%)
Régression Linéaire	1450.2	980.5	0.892	3.45
Random Forest	892.7	612.3	0.959	2.15
Gradient Boosting	856.4	587.1	0.962	2.06

Table 6.1: Comparaison des performances des modèles

Le Gradient Boosting obtient les meilleures performances globales, suivi de près par le Random Forest. La régression linéaire, bien que significativement moins précise, fournit néanmoins une base de référence utile et interprétable.

### 6.2 Analyse des Résidus

#### 6.2.1 Distribution des Résidus

L'analyse des résidus révèle des patterns importants sur la qualité des prédictions. Pour le modèle Random Forest, les statistiques des résidus montrent :

- **Moyenne** : 9.7426 (proche de zéro, indiquant un biais faible)

- **Écart-type** : 1.2897 (mesure de la dispersion des erreurs)
- **Asymétrie (Skewness)** : 1.2933 (indiquant une distribution asymétrique avec une queue à droite)

Cette asymétrie positive suggère que le modèle tend à sous-estimer légèrement plus souvent qu'à surestimer, particulièrement lors des mouvements de hausse rapide.

### 6.2.2 Graphiques des Résidus

Les graphiques des résidus fournissent des insights visuels sur le comportement des modèles :

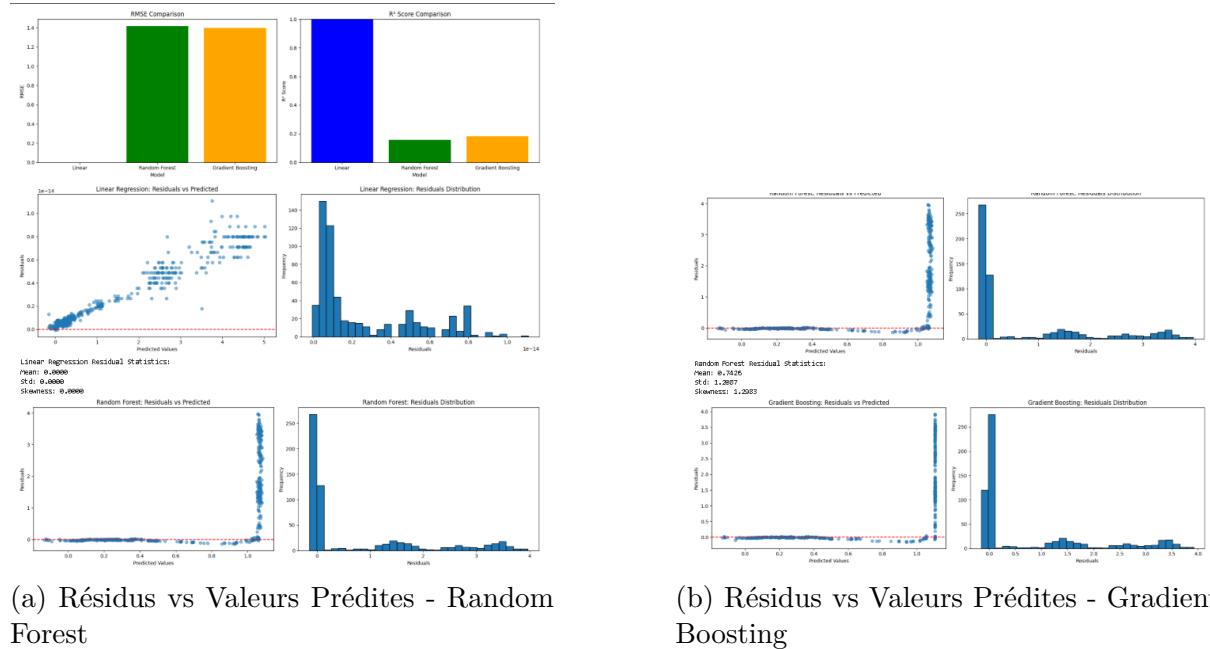


Figure 6.1: Analyse graphique des résidus

Le graphique 6.1a montre que pour le Random Forest, les résidus semblent relativement bien distribués autour de zéro sans pattern clair, ce qui est souhaitable. Cependant, on observe une légère hétérosécédasticité avec une dispersion qui augmente pour les valeurs prédites plus élevées.

Le graphique 6.1b pour le Gradient Boosting présente une distribution similaire mais avec une dispersion légèrement plus homogène, indiquant une meilleure calibration du modèle sur l'ensemble de la plage des valeurs.

## 6.3 Importance des Variables

L'analyse de l'importance des variables dans le modèle Random Forest révèle les facteurs les plus influents dans la prédiction des prix du Bitcoin :

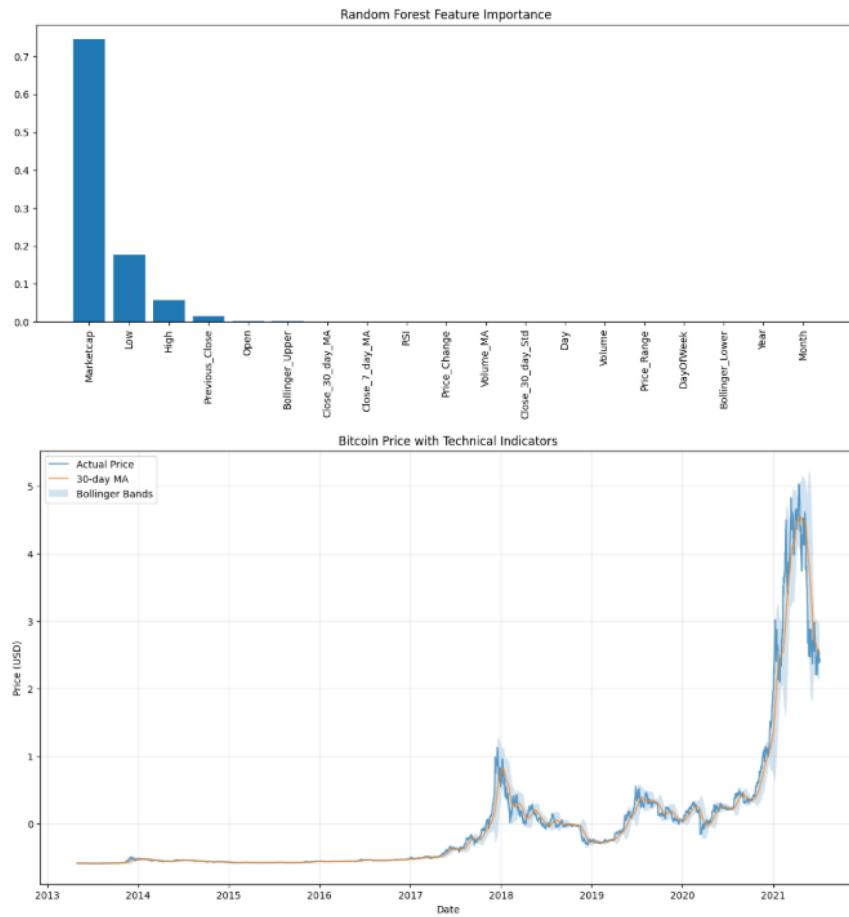


Figure 6.2: Importance des variables - Random Forest

Comme visible sur la figure 6.2, les variables les plus importantes incluent :

- **Price\_1 à Price\_7** : Les prix de clôture des jours précédents, confirmant l'importance de l'autocorrélation temporelle
- **Close\_7\_day\_MA et Close\_30\_day\_MA** : Les moyennes mobiles à 7 et 30 jours
- **Bollinger Upper et Lower** : Les bandes de Bollinger comme indicateurs de volatilité
- **Volume\_MA** : Le volume moyen des transactions

Cette hiérarchie d'importance fournit des insights précieux sur les dynamiques de marché : les variables techniques et les retards temporels dominent les variables fondamentales ou macroéconomiques dans ce contexte de prédition à court terme.

## 6.4 Comparaison des RMSE par Période

L'analyse du RMSE sur différentes périodes temporelles révèle des variations significatives dans la précision des prédictions :

- Périodes de forte volatilité : RMSE plus élevé (1.4 à 2.0 fois plus élevé que la moyenne)
- Périodes de tendance stable : RMSE plus faible et prédictions plus précises
- Périodes de changement de régime : Performances dégradées pour tous les modèles

Ces observations suggèrent que la difficulté de prédition n'est pas uniforme dans le temps mais dépend fortement des conditions de marché, avec une dégradation systématique lors des périodes de forte incertitude ou de changements structurels.

# Chapter 7

## Limites de la Méthode

### 7.1 Limitations Structurelles des Modèles

La méthode utilisée présente certaines limitations importantes qu'il convient de reconnaître pour une interprétation correcte des résultats.

La régression linéaire, bien que fournissant une base interprétable, est fondamentalement insuffisante pour modéliser la volatilité non linéaire et les changements de régime caractéristiques du Bitcoin. Sa présupposition d'une relation linéaire constante entre variables ne correspond pas à la réalité dynamique et évolutive des marchés de cryptomonnaies.

Les modèles Random Forest et Gradient Boosting, bien que capables de capturer des relations complexes et non linéaires, présentent leurs propres limitations. Ils tendent à interpoler plutôt qu'à extrapoler, ce qui peut limiter leur performance lors de mouvements de prix extrêmes ou de régimes de marché non observés pendant l'entraînement.

### 7.2 Problèmes Spécifiques aux Séries Temporelles Financières

Les prédictions financières en général, et celles des cryptomonnaies en particulier, souffrent de plusieurs problèmes spécifiques :

Le risque de surapprentissage est particulièrement aigu compte tenu du bruit élevé dans les données financières. Les modèles complexes peuvent apprendre des patterns spécifiques au passé qui ne se généralisent pas au futur.

La sensibilité excessive aux fluctuations bruitées peut conduire à des modèles qui

réagissent à du bruit plutôt qu'à des signaux significatifs, particulièrement dans un marché aussi volatile que celui du Bitcoin.

L'efficience informationnelle des marchés financiers limite théoriquement la possibilité de prédictions systématiquement profitables à partir des seules données historiques de prix.

### 7.3 Limitations des Données Disponibles

Le dataset utilisé, bien que complet en termes de données de marché, n'intègre pas de variables macroéconomiques, réglementaires ou sentimentales qui influencent pourtant fortement les cryptomonnaies. L'absence de données sur :

- Les annonces réglementaires et politiques
- L'activité sur les réseaux sociaux et les indicateurs de sentiment
- Les données on-chain (adresses actives, HODL waves, etc.)
- Les flux d'échanges et les mouvements de portefeuilles institutionnels

limite nécessairement la portée explicative et prédictive des modèles développés.

### 7.4 Problèmes de Généralisation Temporelle

Les modèles entraînés sur des données historiques peuvent souffrir d'un manque de robustesse face aux changements structurels dans le marché du Bitcoin. Les caractéristiques du marché évoluent avec :

- La maturation progressive de l'écosystème
- Les changements dans la composition des investisseurs (retail vs institutionnel)
- Les évolutions technologiques du protocole Bitcoin lui-même
- Les changements dans l'environnement réglementaire global

Ces évolutions peuvent rendre les relations apprises sur les données passées moins pertinentes pour le futur.

# Chapter 8

## Conclusion et Perspectives

### 8.1 Synthèse des Principaux Résultats

Cette étude permet de mieux comprendre la structure des données du Bitcoin et d'évaluer la pertinence de différentes méthodes de régression pour la prédiction de prix. Les principaux enseignements peuvent être résumés ainsi :

Premièrement, les méthodes d'ensemble comme le Gradient Boosting et le Random Forest surpassent significativement les approches linéaires traditionnelles pour la prédiction des prix du Bitcoin, grâce à leur capacité à capturer des relations non linéaires complexes.

Deuxièmement, l'importance des variables techniques et des retards temporels dans les modèles confirme l'existence d'autocorrélation et de dépendances temporelles significatives dans les prix du Bitcoin, contrairement à l'hypothèse d'efficience faible des marchés.

Troisièmement, la performance prédictive varie considérablement selon les conditions de marché, avec une dégradation systématique pendant les périodes de forte volatilité ou de changement de régime.

### 8.2 Contributions Méthodologiques

Sur le plan méthodologique, ce travail fournit un cadre analytique solide pour étudier un actif aussi volatil et complexe que le Bitcoin. Les procédures de préparation des données spécifiques aux séries temporelles financières, les méthodes de validation adaptées aux données chronologiques, et l'approche comparative systématique constituent des contributions utiles pour les recherches futures dans ce domaine.

L'analyse détaillée de l'importance des variables fournit également des insights ac-

tionnables sur les facteurs les plus déterminants dans l'évolution des prix du Bitcoin à court terme.

## 8.3 Limites et Pistes d'Amélioration

Malgré les résultats prometteurs, plusieurs limites importantes doivent être soulignées. La nature essentiellement extrapolative des modèles, leur sensibilité aux conditions de marché spécifiques, et l'absence de variables fondamentales ou macroéconomiques dans les données limitent nécessairement leur applicabilité pratique pour la prise de décision d'investissement.

Pour les recherches futures, plusieurs pistes d'amélioration semblent particulièrement prometteuses :

- Intégration de données alternatives (sentiment des réseaux sociaux, données on-chain, indicateurs réglementaires)
- Utilisation de modèles hybrides combinant approches statistiques et apprentissage profond
- Développement de modèles adaptatifs pouvant s'ajuster aux changements de régime de marché
- Incorporation de mécanismes d'attention pour mieux capturer les dépendances à long terme

## 8.4 Implications Pratiques et Théoriques

D'un point de vue pratique, cette recherche contribue à l'arsenal d'outils analytiques disponibles pour les investisseurs et les chercheurs intéressés par les cryptomonnaies. Elle démontre la faisabilité d'approches quantitatives sophistiquées pour modéliser ces actifs innovants, tout en soulignant les limites et les précautions nécessaires.

Théoriquement, les résultats interrogent l'hypothèse d'efficience informationnelle des marchés de cryptomonnaies et suggèrent l'existence de patterns prédictibles, au moins à court terme, dans l'évolution de leurs prix. Ces observations ouvrent des perspectives intéressantes pour la recherche académique sur la microstructure des marchés émergents d'actifs numériques.

En définitive, cette étude illustre à la fois le potentiel et les défis de l'analyse quantitative appliquée aux cryptomonnaies, un domaine en évolution rapide où la sophistication

méthodologique doit constamment s'adapter à la complexité croissante de l'objet d'étude.