

Script Présentation TIPE

1. Peut-on prédire les cours des cryptomonnaies ?

- Étude des cours des cryptomonnaies afin d'en prévoir les tendances

2. Qu'est ce qu'une cryptomonnaie ?

- Devise (comme le dollar ou l'euro) mais décentralisée (qui ne dépend pas de banque centrale)
- Transactions inscrites sur un registre virtuel appelé blockchain
- Connue pour avoir des cours très volatiles et spéculatifs (idées de profits très rapides mais réalité plus compliquée)
- Les deux cryptomonnaies les plus connues sont le Bitcoin et l'Ethereum

3. Présentation du projet

- Création d'une base de donnée à partir des cours historiques des cryptomonnaies
- Analyse des cours grâce à des algorithmes de reconnaissance de formes afin de formuler des lois empiriques pour en prévoir les tendances
- Implémentation des modèles grâce à une seconde base de données indépendante

4. Création des bases de données

- Deux bases de données : la première, plus grosse, regroupe les cours de six cryptomonnaies de 2017 à 2022 (plus de dix millions de valeurs). La deuxième est constituée de données plus récentes et permettra de valider ou non les différents modèles
- Données récupérées grâce au module Python CCXT : module utilisé en finance, permet de se connecter aux grandes plateformes d'échange de cryptomonnaies (Binance) et ici de récupérer les valeurs des cours historiques des cryptomonnaies
- Les valeurs sont stockées dans des fichiers texte afin d'en faciliter la manipulation

5. Organisation des données

- Données correspondent aux valeurs des cours par minute
- Organisation sous formes de listes de 60 valeurs → une liste correspond à une plage d'une heure
- Établissement d'une liste : échantillonnage du cours toutes les minutes

6. Étude de la corrélation des cours

- Étude de la corrélation des cours afin de simplifier le problème et d'alléger les temps de comparaison
- Cours positivement corrélés (coefficient de corrélation moyen de 0,82)
- Exemple de deux cryptomonnaies corrélées
- Cryptomonnaies corrélées au Bitcoin → Simplification du problème à l'étude du cours du Bitcoin

7. Première modélisation : Codage sur un bit

- Première modélisation : coder les valeurs sur 1 bit correspondant aux variations du cours
- 0 si le cours est en baisse, 1 s'il est en hausse → la modélisation ne prend en compte uniquement que les variations locales du cours
- Variations de listes de 60 valeurs → Listes de 59 bits
- Exemple concret d'implémentation

8. Comparaison des listes : Distance de Hamming

- Les listes sont comparées avec la distance de Hamming
- Définition de la distance de Hamming : le nombre d'éléments de deux suites binaires qui diffèrent, exemple
- Concrètement, on recherche les dix listes les plus proches du cours évalué dans la base de données au sens de la distance de Hamming, puis on compare les tendances sur l'heure suivante

9. Définition d'un « succès »

- Il y a succès lorsque la tendance prédite se réalise
- Le cours en bleu correspond au cours dont on veut prédire la tendance
- Le cours en gris correspond à une prédiction (de la base de données) dont l'évolution est connue
- Le cours en rouge correspond à l'évolution réelle du cours évalué
- À gauche, la prédiction n'a pas su prévoir la tendance du cours, c'est un échec
- À droite, l'évolution réelle suit la tendance de la prédiction, c'est un succès

10. Résultats de la première modélisation

- La première modélisation ne donne pas de résultats satisfaisants : 49,5 % de succès, moins de 1/2 → modélisation inutilisable
- Les prédictions être « proches » du cours évalué sur l'heure de comparaison ne reproduisent même pas sa tendance
- Modélisation trop locale, nécessité de tenir compte de la tendance globale de la liste

11. Deuxième modélisation : Codage sur 8 bits

- Deuxième modélisation : afin de tenir compte de la tendance globale de la liste, on code les valeurs sur 8 bits
- La plus petite valeur de la liste se voit assigner la valeur 0, la plus haute la valeur 255 et les autres proportionnellement
- Listes de 60 valeurs entières entre 0 et 255

12. Résultats de la deuxième modélisation

- Les prédictions suivent la tendance du cours évalué sur la première heure, contrairement à la première modélisation
- Toutefois, cela ne semble pas garantir que ceux-ci prévoient la tendance de l'évolution réelle du cours → modélisation encore chaotique

- Taux de succès de 51 %, reste faible mais supérieur à 1/2 donc encourageant → incitation à se concentrer sur les tendances globales des listes et non leurs petites variations

13. Un résultat intéressant

- Comparaison des tendances de deux listes successives
- Si la tendance du cours a été haussière au cours de la dernière heure, elle sera probablement baissière au cours de l'heure suivante et inversement
- Plus forte probabilité que le cours change de tendance qu'il ne garde la même

14. Troisième modélisation

- Troisième modélisation dans le sens de l'observation précédente
- On repère la tendance (haussière ou baissière) puis on repère la position de l'extremum
- Établissement des lois empiriques en fonction de la tendance et de la position de l'extremum :
le graphe montre la probabilité, selon la tendance et la position de l'extremum, que la tendance de l'heure suivante soit haussière
- Si l'extremum se réalise vers la fin de la liste (donc récemment), il y a de fortes probabilités que le cours change de tendance

15. Résultats de la troisième modélisation

- Exemple d'un cours en baisse et dont l'extremum a été atteint en fin de liste → conformément à la loi énoncée, le cours change de tendance
- La modélisation renvoie la bonne tendance dans 53 % des cas, ce qui permet d'avoir une longueur d'avance sur l'investisseur moyen (qui fait souvent des choix irrationnels, donc réussit son pari dans 50 % des cas voire pire) → modélisation exploitable

16. Implémentation dans le cas réel

- Imaginons que l'on investisse dans un bitcoin le 16 janvier. Deux possibilités s'offrent à nous :
- Première possibilité, on laisse le bitcoin en l'état sans y toucher. Celui-ci a perdu 33 % de sa valeur au 29 mai donc notre capital s'est déprécié d'un tiers
- Seconde possibilité, on réévalue notre investissement toutes les heures suivant la modélisation décrite précédemment. L'algorithme réussit à avoir un retour sur investissement de 8 % sur 4 mois soit un retour sur investissement moyen de 26 % par an, performance tout à fait honorable d'autant plus que le marché est dans une phase baissière

Conclusion

Si la prédiction des tendances des cours à moyen/long terme relève plutôt de l'analyse fondamentale et une fine connaissance des marchés, il est possible de formuler des lois et de dégager des tendances à court terme permettant d'optimiser ses gains et de lisser les risques, en réussissant à dégager un profit alors même que les cours sont en baisse.