

PROJET :JOUONS ET GAGNONS EN BOURSE

Contexte et Objectifs

Bonjour à toutes et à tous,

Ces dernières années, l'investissement en bourse est devenu une tendance populaire, notamment avec l'essor du **trading en ligne** facilité par des plateformes accessibles et des tutoriels YouTube. Beaucoup rêvent de suivre les traces de **Warren Buffett**, convaincus qu'avec les bonnes stratégies, ils pourront multiplier leur capital et atteindre l'indépendance financière.

Cependant, le trading et l'investissement en bourse ne se résument pas à suivre des intuitions ou des recommandations aléatoires. Une analyse rigoureuse des données financières, une bonne compréhension des modèles économiques et une maîtrise des outils technologiques sont essentielles pour prendre des décisions éclairées.

Ce projet vise à analyser le marché boursier et à développer des modèles de prévision des prix des actions en utilisant des techniques de Data Science et de Machine Learning.

Enjeux et Approche

Vous allez vous mettre dans la peau d'un analyste financier et explorer les actions de différentes entreprises afin de :

- Comprendre le marché et les dynamiques des actions
- Appliquer des techniques d'analyse de données (préprocessing, PCA, clustering)
- Détecter des anomalies sur certaines périodes
- Construire des modèles prédictifs pour anticiper l'évolution des prix des actions
- Évaluer et comparer la performance des modèles

Vous travaillerez sur des données de marché réel, issues des séances précédentes ou d'autres sources en ligne (ex : Yahoo Finance, Boursorama). Votre travail devra être rigoureux, structuré et interprété avec une rédaction professionnelle.

Format attendu du rendu

Votre rendu se fera sous la forme **d'un Notebook Jupyter** contenant :

- **Le code source** bien commenté
- **L'export au format PDF ou HTML** si possible
- **Des analyses détaillées avec des graphiques, des justifications et des interprétations claires**
- **Une présentation en format pptx**

Vous pouvez utiliser les notebooks fournis en cours, mais veuillez à bien les adapter aux besoins du projet. L'objectif est d'améliorer le pouvoir prédictif de vos modèles en appliquant des méthodologies adaptées.

PARTIE A : ANALYSE ET PRÉTRAITEMENT

1. Exploration et reporting

- Effectuez une **analyse exploratoire des données** pour obtenir un premier aperçu.
- Identifiez les caractéristiques clés et repérez les éventuelles valeurs aberrantes.

2. Prétraitement des données

- Normalisez ou standardisez les variables si nécessaire.
- Vérifiez les données manquantes et appliquez des stratégies de gestion adaptées.

4. Segmentation du marché boursier (Clustering)

- Appliquez K-Means avec $k=2$ et interprétez les groupes obtenus.
- Déterminez le nombre optimal de clusters via la méthode du coude et silhouette score.
- Réalisez un clustering hiérarchique et comparez les résultats avec K-Means.
- Visualisez les clusters sur le plan défini par les deux premières composantes principales.

5. Détection d'anomalies

- Identifiez les 5% de jours considérés comme anomalies à l'aide d'un modèle de détection.
- Expliquez les causes potentielles de ces anomalies (événements macroéconomiques, crises, décisions d'entreprise...).

PARTIE B : PRÉVISION DES ACTIONS DE TESLA

1. Analyse des données de Tesla

- Effectuez une analyse statistique descriptive des actions de Tesla.
- Vérifiez si les prix des actions suivent une distribution normale et/ou s'ils sont stationnaires via des tests statistiques (ex : test de Dickey-Fuller).

3. Modélisation avancée avec SARIMAX

- Créez un modèle $ARIMA(p,d,q)$ et analysez ses performances.
- Étendez à un modèle $SARIMA(p,d,q)(P,D,Q,S)$ pour capturer la saisonnalité des prix.

4. Influence des autres actions sur Tesla

- Étudiez si les actions de Facebook, Amazon, Apple, Google et d'autres influencent Tesla via un test de causalité de Granger.
- Entraînez un nouveau modèle SARIMAX en intégrant les actions ayant un impact significatif.

5. Modèles avancés en Deep Learning (Bonus)

- Expérimentez avec Facebook Prophet et entraînez un modèle prédictif.

- Testez des architectures LSTM (Long Short-Term Memory) pour la prévision des séries temporelles.
- Comparez la performance des modèles SARIMAX Prophet et LSTM sur la base de métriques comme RMSE, MAE, R^2 .

PARTIE C : DÉPLOIEMENT DU MODÈLE

1. Préparation du modèle pour le déploiement

Une fois votre modèle entraîné et validé, il est crucial de penser à son mise en production afin qu'il puisse être utilisé pour des prédictions en temps réel ou en batch.

- Choix du modèle final : Quel modèle est le plus performant et interprétable pour un usage en production ?
- Exportation du modèle : Sauvegardez le modèle au format `pickle`, `joblib` ou sous forme d'API pour une utilisation dynamique.
- Optimisation des performances : Y a-t-il un compromis entre précision et rapidité d'exécution ?

2. Création d'une API de prédiction

L'objectif est de permettre à des utilisateurs d'accéder aux prédictions via une API.

- Implémentez une API avec Flask ou FastAPI ou streamlit pour interroger votre modèle.
- Soumettez de nouvelles données afin d'obtenir des prévisions.

QUESTIONS OUVERTES POUR AFFINER LES PRÉVISIONS (bonus)

Dans un contexte de prévision boursière avancée, plusieurs facteurs externes peuvent influencer les résultats. Il est essentiel de se poser les bonnes questions et d'intégrer des régresseurs additionnels pour améliorer la précision du modèle.

1. Largeur des régresseurs et variables explicatives

- Quelles autres variables pourraient améliorer nos prévisions ?
 - Indicateurs macroéconomiques (inflation, taux d'intérêt, chômage)
 - Cours d'autres actions corrélées (ex : Orange influence Tesla ?)
 - Volume des transactions
 - Sentiment analysé via des articles ou tweets financiers
- Comment tester l'impact de ces variables sur notre modèle ? (Feature importance, Granger Causality Test)

2. Événements et jours fériés (event, holiday)

- Faut-il prendre en compte des événements spécifiques dans nos prévisions ?
 - Résultats trimestriels des entreprises
 - Annonce de nouvelles technologies (Tesla, Apple, Nvidia)
 - Conférences financières majeures (ex : FED, BCE)
 - Changements de réglementation
- Quels sont les jours fériés ou périodes de fermeture du marché qui influencent la liquidité et la volatilité ?

3. Cycles et tendances long terme

- Y a-t-il des cycles récurrents dans l'évolution des prix (effet janvier, bull/bear market) ?
- Comment détecter les tendances à long terme et distinguer les effets saisonniers des tendances structurelles ?
- Quelle est la périodicité optimale pour lissage exponentiel ou SARIMA ?

4. Variables connues dans le futur (Future known regressors)

Certaines variables sont connues avant la prédiction et peuvent être intégrées dans nos modèles.

- Quels indicateurs sont prédictibles avec certitude ? (taux d'intérêt prévus, dividendes annoncés...)
- Comment intégrer ces variables dans des modèles comme Facebook Prophet ou LSTM ?

Si vous avez des questions ou besoin d'une orientation sur l'approche méthodologique, n'hésitez pas à me contacter.

Bon courage !

Mr. SAMBE