

Travaux Dirigés Notés – 3A_DDEFI : Prédiction des Prix et Actifs du S&P 500 grâce à du Machine Learning simple

Contexte :

Les marchés financiers sont complexes et influencés par de nombreux facteurs économiques, techniques et comportementaux. Dans ce TD, vous allez développer un projet de machine learning pour prédire les prix et les rendements des actifs du **S&P 500**. Ce projet mettra en application des techniques avancées d'analyse de séries temporelles, de modélisation prédictive et d'évaluation des performances. Vous vous aiderez des conseils vus en présentiel lors des [cours](#).

Objectifs du Projet :

1. **Collecter et préparer les données financières** (prix historiques, indicateurs macroéconomiques).
 2. **Développer un modèle prédictif** en utilisant des algorithmes de machine learning adaptés.
 3. **Évaluer les performances** de votre modèle avec des métriques appropriées.
 4. **Interpréter les résultats** pour proposer des recommandations d'investissement.
-

Partie 1 : Collecte et Préparation des Données

1. **Collecte des données (+++) :**
 - Récupérez les données historiques du S&P 500 (prix de clôture, volume) sur une période de 5 ans via une API (Yahoo Finance ou Alpha Vantage).
 - Ajoutez des variables exogènes (VIX, taux d'intérêt, indicateurs économiques).
2. **Prétraitement des données (++) :**
 - Nettoyez les données (valeurs manquantes, doublons).
 - Transformez les prix en rendements log (log-returns).
 - Effectuez une analyse de stationnarité (test ADF) et déterminez les décalages temporels pertinents.
3. **Feature engineering (+) :**
 - Créez des indicateurs techniques (Moyennes mobiles, RSI, MACD).
 - Identifiez les variables clés avec des techniques de corrélation ou de sélection de features.

Partie 2 : Développement du Modèle Prédictif

1. Choix des modèles :

- Implémentez **au moins deux modèles** parmi les suivants :
 - **Régression linéaire** pour les rendements. +++
 - **Modèle ARIMA** pour les séries temporelles stationnaires.
 - **Random Forest** ou **Gradient Boosting** pour la prédiction non linéaire.
 - **LSTM (Long Short-Term Memory)** pour capturer les dépendances temporelles (quasi optionnel).

2. Entraînement et validation :

- Séparez les données en ensemble d'entraînement (80%) et de test (20%) cf Pareto.
- Utilisez une **validation croisée temporelle** (rolling window).
- Évaluez les modèles avec les métriques suivantes :
 - **MAE (Mean Absolute Error)**
 - **RMSE (Root Mean Squared Error)**
 - **Accuracy** si vous prédisiez les directions (hausse/baisse).

Partie 3 : Analyse des Résultats et Interprétation

1. Analyse des performances :

- Comparez les performances des modèles.
- Interprétez les résultats à l'aide de graphiques (évolution des erreurs, prédictions vs valeurs réelles).

2. Discussion sur les erreurs courantes :

- Identifiez les sources potentielles d'erreur (surajustement, multicolinéarité, endogénéité).
- Proposez des pistes d'amélioration.

3. Conclusion et recommandations :

- Présentez les principales conclusions de votre projet.
- Proposez des stratégies d'investissement basées sur vos prédictions.

Livrables Attendus sur le Github du groupe :

1. **Rapport écrit** (10 pages max) comprenant :
 - Introduction et objectifs.
 - Méthodologie (préparation des données, choix des modèles).
 - Résultats et analyse.
 - Conclusion et recommandations.
2. **Code Python** documenté et reproductible.
3. **Support Visuel** (10 slides max) : synthèse du projet, avec support visuel.

Critères d'Évaluation :

- **Qualité de la collecte et de la préparation des données** (30%).
- **Rigueur de la modélisation et choix des algorithmes** (20%).
- **Analyse des résultats et interprétation des erreurs** (20%).
- **Clarté du rapport** (30%).

Ressources :

- Documentation Python : [pandas](#), [scikit-learn](#), [statsmodels](#), [tensorflow](#).
- Base de données : Yahoo Finance, Alpha Vantage API.
- Articles de référence sur la prédiction boursière avec le machine learning++

Bon courage et bonne analyse !

Le nombre d'heures de cours étant limité, le projet attendu pour le mid term n'est pas un rapport 'parfait' mais une structure claire et cohérente.

L'objectif principal est que vous commenciez à vous coordonner efficacement pour poser les bases solides de votre projet final qui vous servira à étoffer vos portfolios.

N'hésitez pas à [nous contacter](#) pour toute question.

Début: 03.12.2024

Deadline: 07.01.2024

Lirone & Sitraka