

PROJET IA

Commodity Market

Prédiction du cours du blé

Mouthiez François
Kabrit Matteo
Rollet Gabriel
Valletoux Baptiste



Sommaire

1- Introduction

2- Présentation de la data

4- Modèles de ML utilisés, choix du modèle et optimisation

5- Résultats et limites de notre projet

6- Conclusion

Introduction

Dans un monde où les marchés agricoles sont soumis à une volatilité croissante, la capacité à prédire avec précision les fluctuations des prix du blé est d'une importance capitale. Cette initiative trouve son utilité dans une multitude de domaines, offrant des avantages significatifs pour divers acteurs économiques :

Les producteurs agricoles

- Optimiser leurs décisions d'investissement
- Réduire les risques financiers
- Soutenir la durabilité agricole

Les marchés financier

- Utiliser ces prévisions pour élaborer des stratégies d'arbitrage
- le développement de nouveaux produits financiers innovants (des ETF agricoles, des produits structurés basés sur les performances des marchés des céréales

La sécurité alimentaire mondiale

- Meilleure gestion des crises alimentaires
- Assurer un approvisionnement stable en céréales pour les populations vulnérables.

Présentation de la data

#	Column	Non-Null Count	Dtype
0	date	2528 non-null	datetime64[ns]
1	balticdryindex	2446 non-null	float64
2	GPRindex	2524 non-null	float64
3	USDEUR	2528 non-null	float64
4	BRENT	2453 non-null	float64
5	OVXcrudeoil	2527 non-null	float64
6	SPGSCI	2528 non-null	float64

Data allant du 31/03/2014 au 12/04/2024

- Données non journalières, basées seulement sur les journées d'ouverture de marché du S&P GSCI Wheat Index.
- Nécessité de gérer la data manquante

```
82 lignes manquantes.  
2528 lignes au total.  
3.2436708860759493 % de données manquantes.
```

Nécessité de faire un feature scaling

- Normalisation

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

[0 ; 1]

Variables explicatives (features)

2 variables extra-financières:

Baltic Dry Index : indice des prix pour le transport maritime de vrac sec (principalement minerais, charbon et céréales).

GPR Index : mesure des événements géopolitiques défavorables et des risques associés sur la base d'un décompte des articles de presse traitant des tensions géopolitiques.

4 variables financières:

BRENT
USD/EUR
OVX Crude Oil



Notre target

Commodities
S&P GSCI WHEAT

The S&P GSCI Wheat Index, a sub-index of the S&P GSCI, provides investors with a reliable and publicly available benchmark for investment performance in the wheat commodity market.

Justification du choix des features

Intuitions:

Baltic Dry Index

Mécanisme d'offre et de demande:

- Les prix de transport maritime de matières premières peuvent refléter le **surplus/pénurie**.
- Sert de référence pour les systèmes **d'assurance/couverture de risques**.

GPR Index

Appréciation de la **volatilité**:

- Permet de déceler les **risques** planant sur les exploitations de blé à l'échelle mondiale.
- Un **risque géopolitique élevé** laisse présager une **baisse des investissements, des cours boursiers et de l'emploi**.

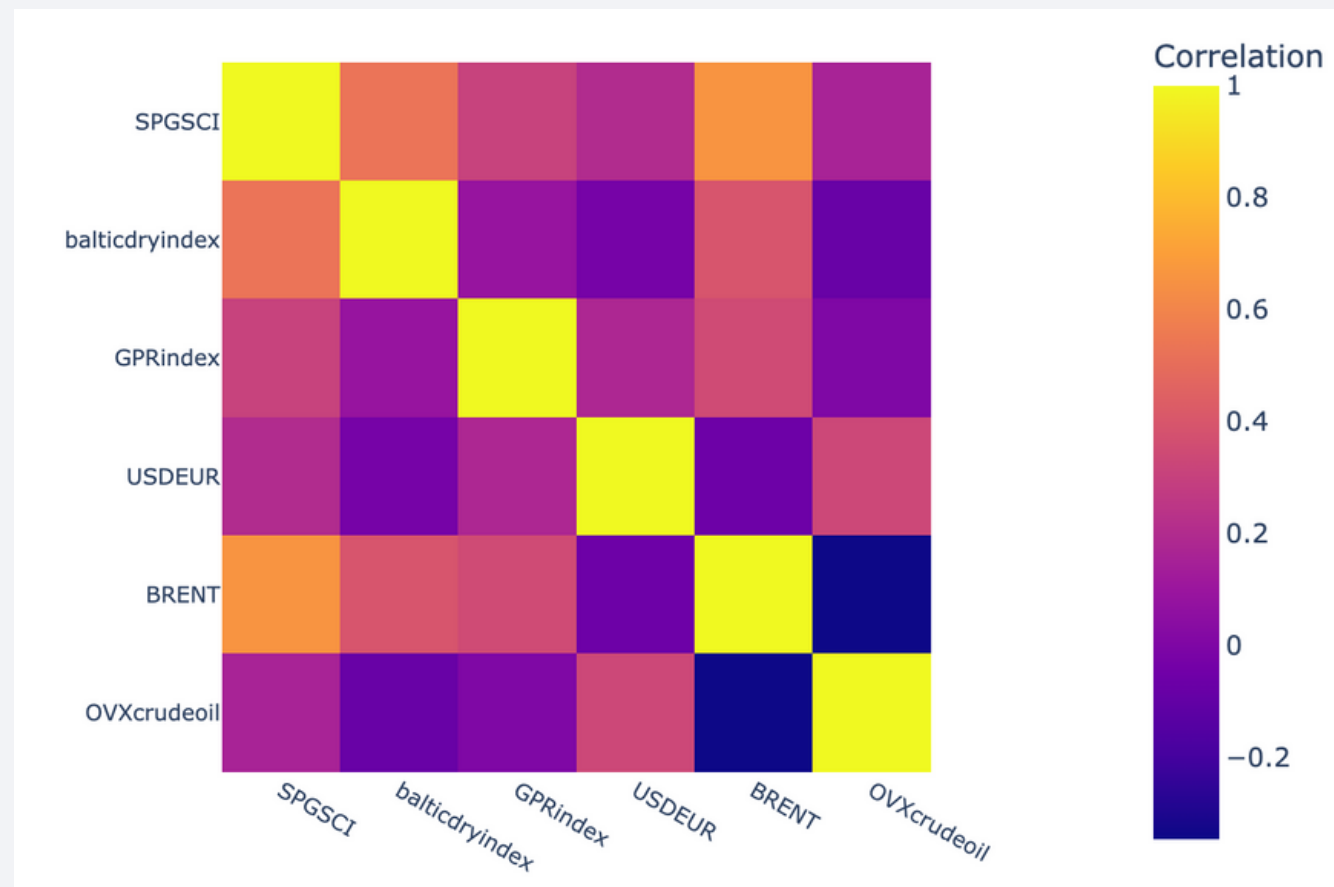
BRENT & OVX Crude oil

- Une variation du cours du Brent impact les coûts de production agricole.
- Le marché pétrolier est très largement surveillé par les investisseurs, impactant de nombreux secteurs.

USD/EUR

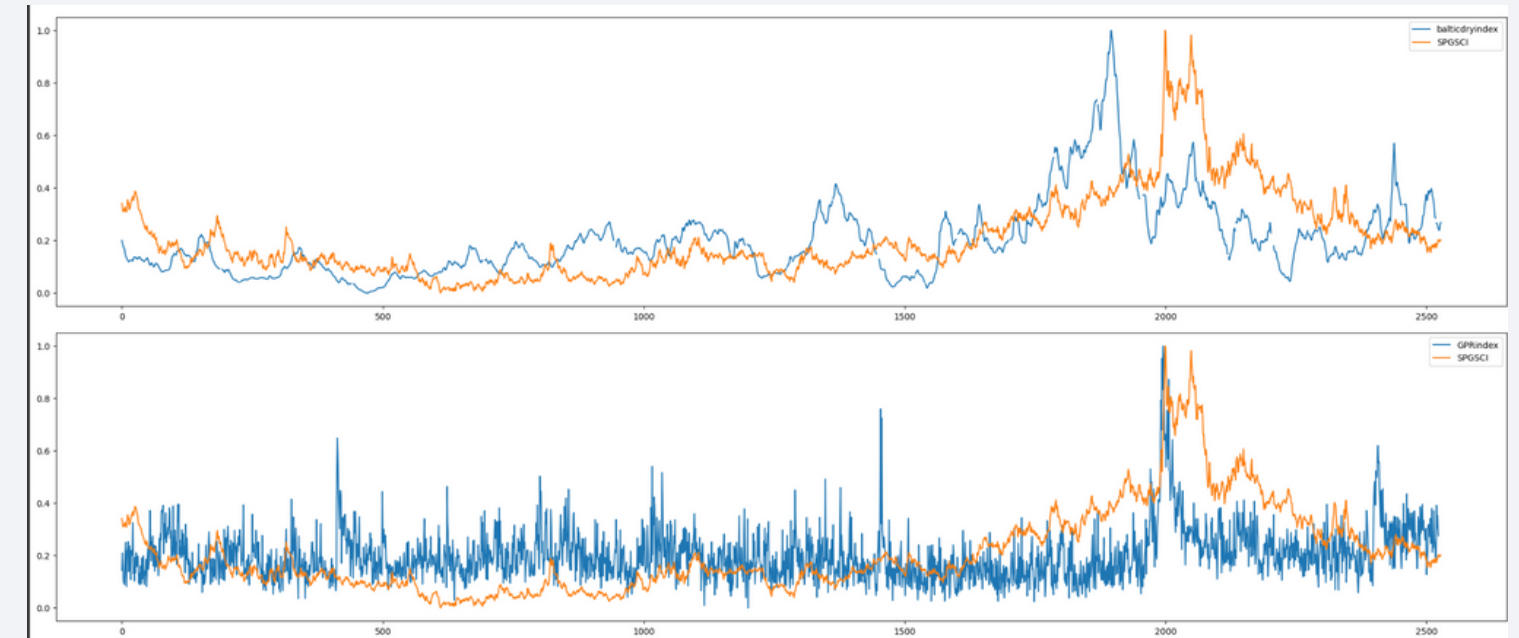
- Les États-Unis et l'Union européenne sont **deux des principaux producteurs et exportateurs mondiaux** de blé.
- Les marchés à terme et les bourses de matières premières les plus influentes, telles que le **Chicago Mercantile Exchange (CME)** où sont négociés les contrats à terme sur le blé, sont basées aux États-Unis.

Analyse de la data



- Etude des corrélations entre les variables et de leur évolution historique avec notre target.
- Première intuition:
 - Les variables semblent bien expliquer notre target, sans pour autant présenter une évolution identique à celle de notre target.

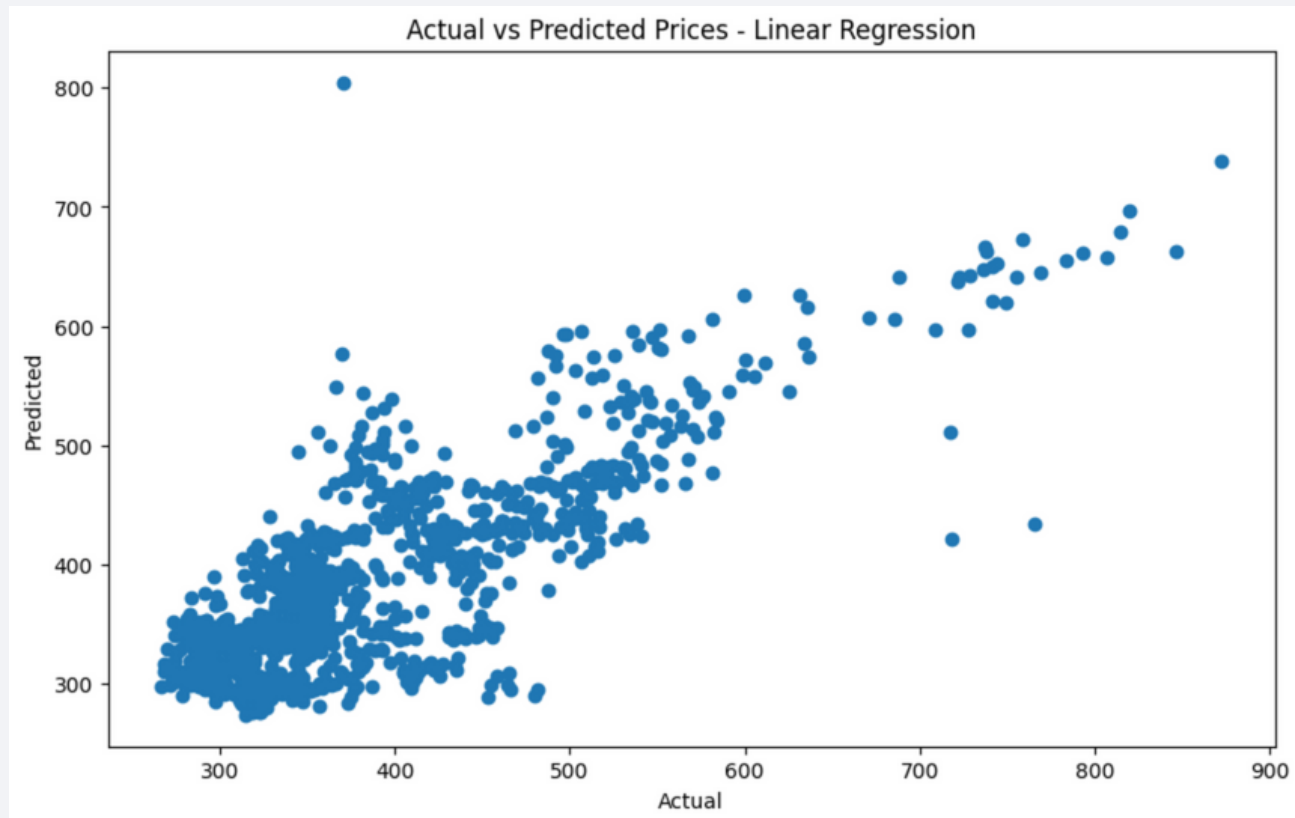
- Baltic dry index et GPR Index



- USD EUR, Brent et OVX Crude Oil

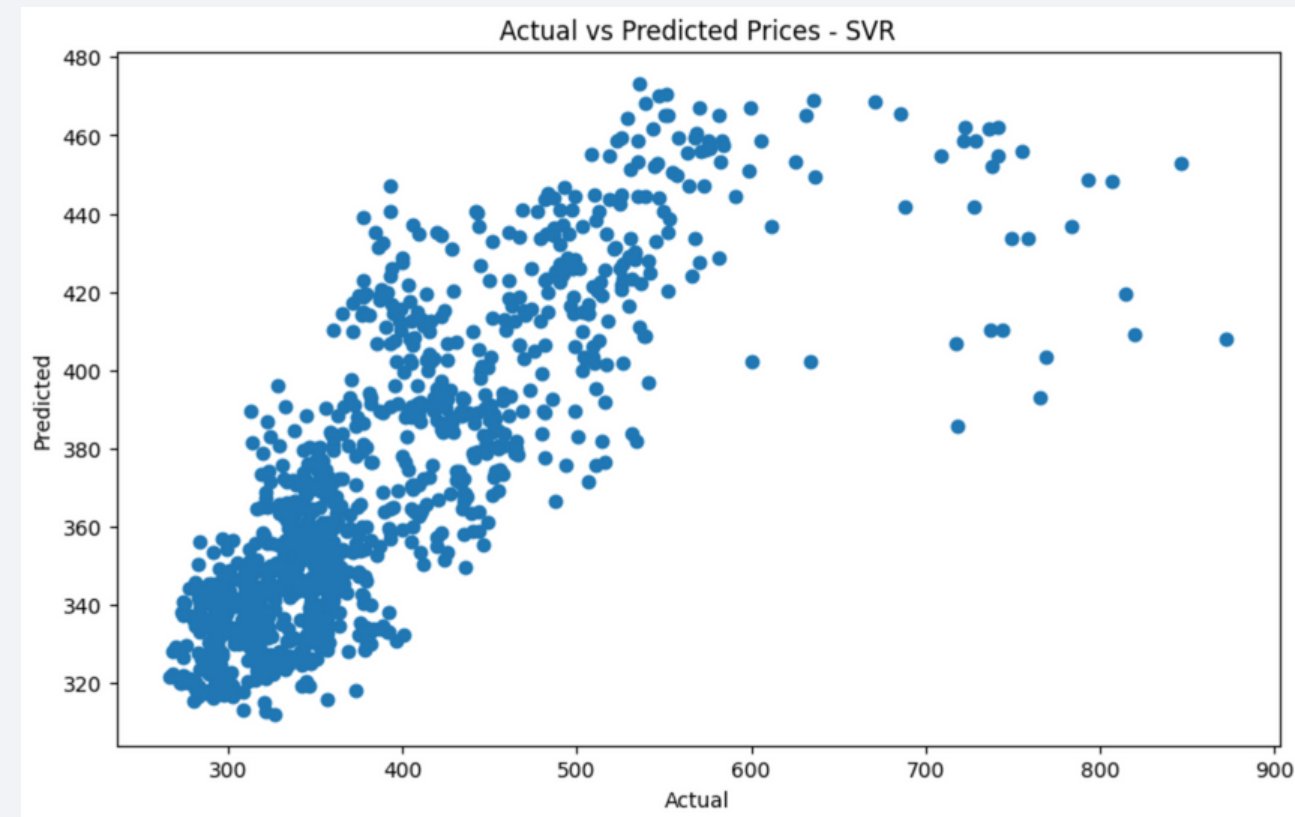
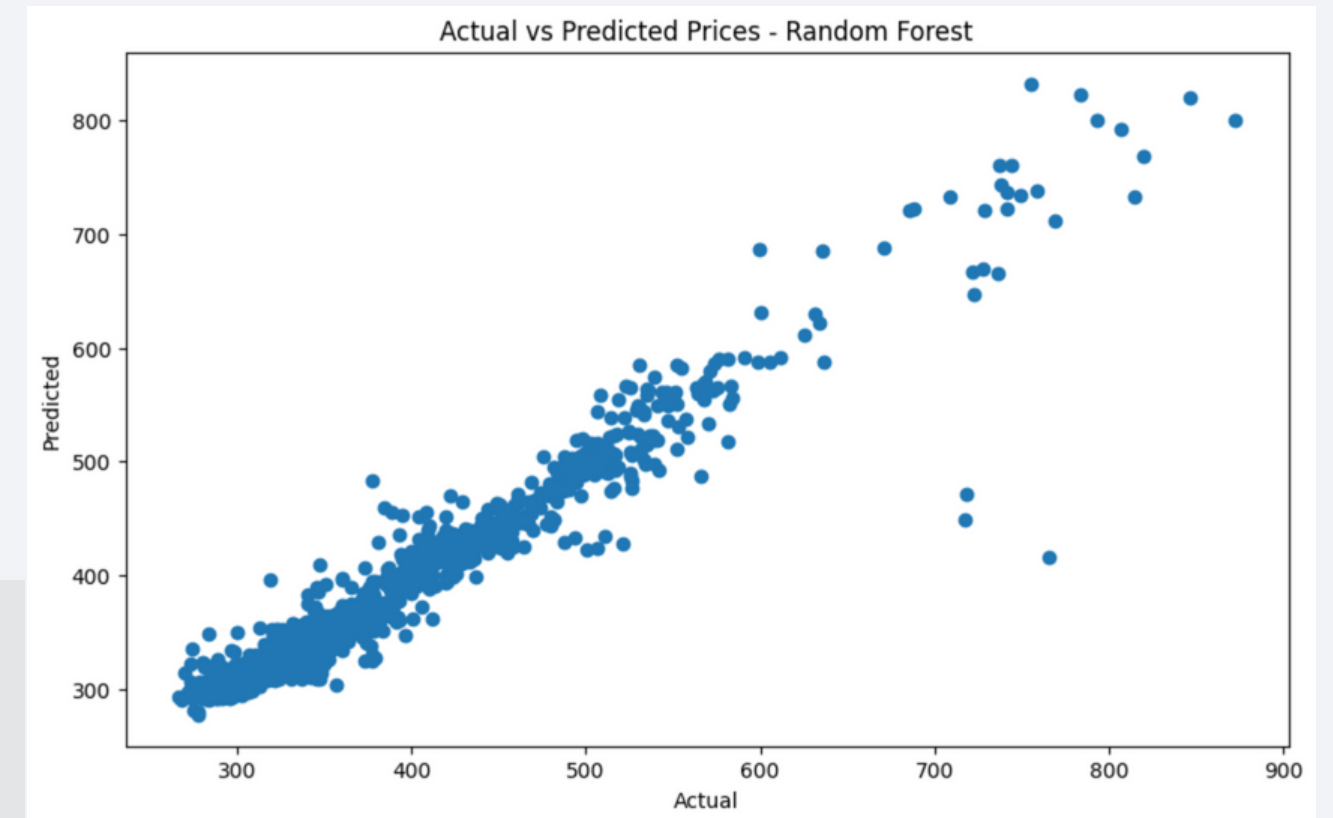


Modèles de ML utilisés



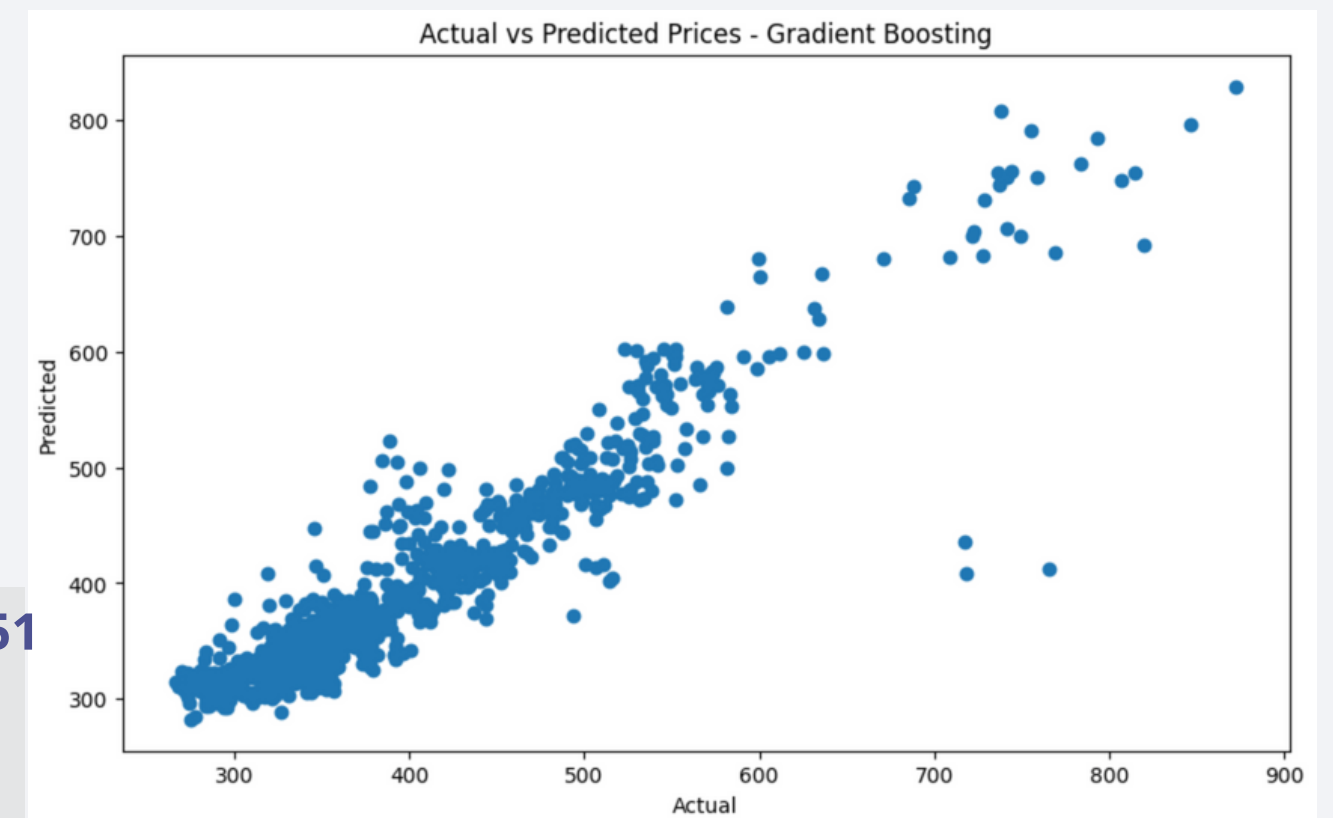
Mean Squared Error: 3596.16
Mean Absolute Error: 46.10
R-squared: 0.63

Mean Squared Error: 661.75
Mean Absolute Error: 15.00
R-squared: 0.93

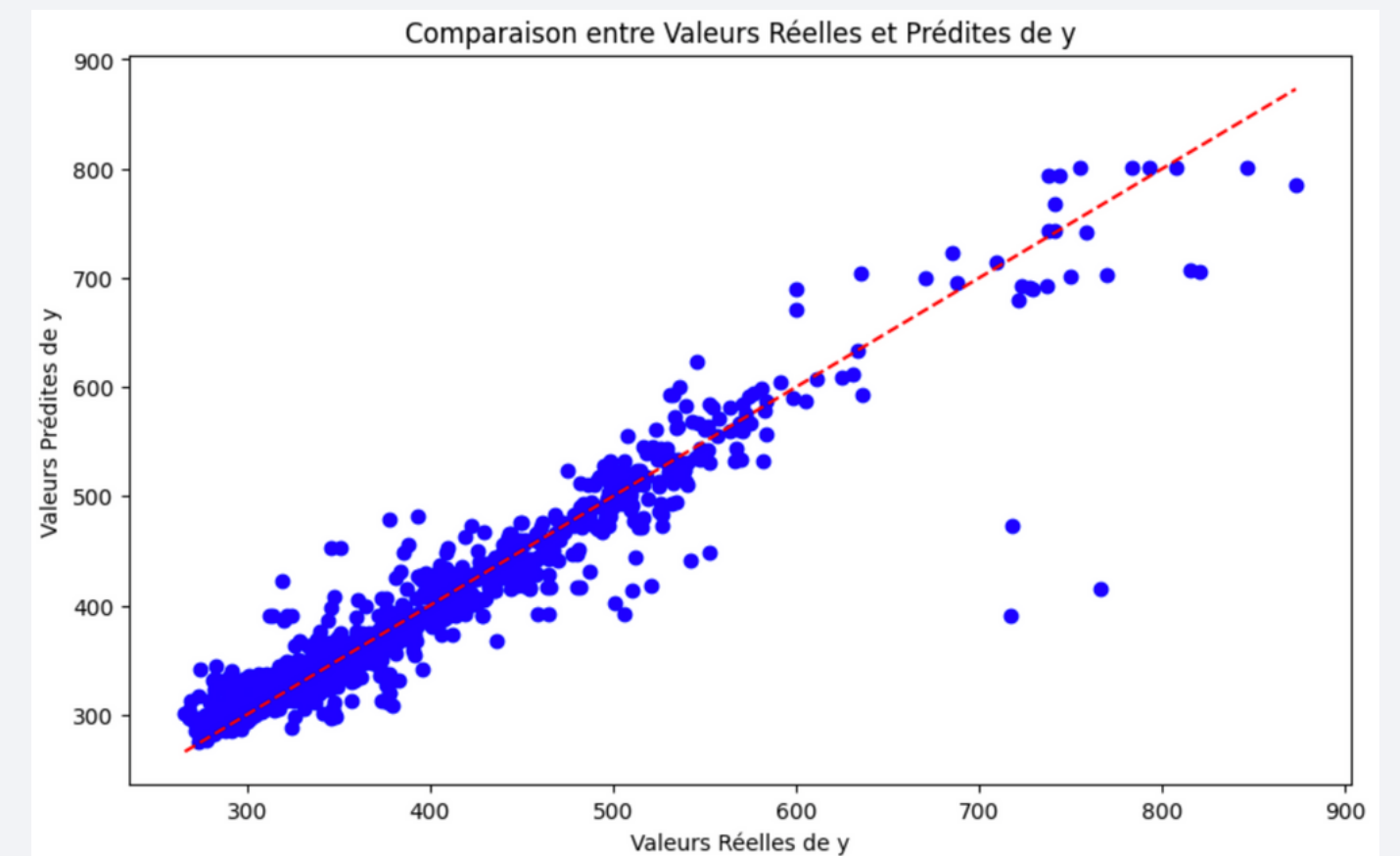
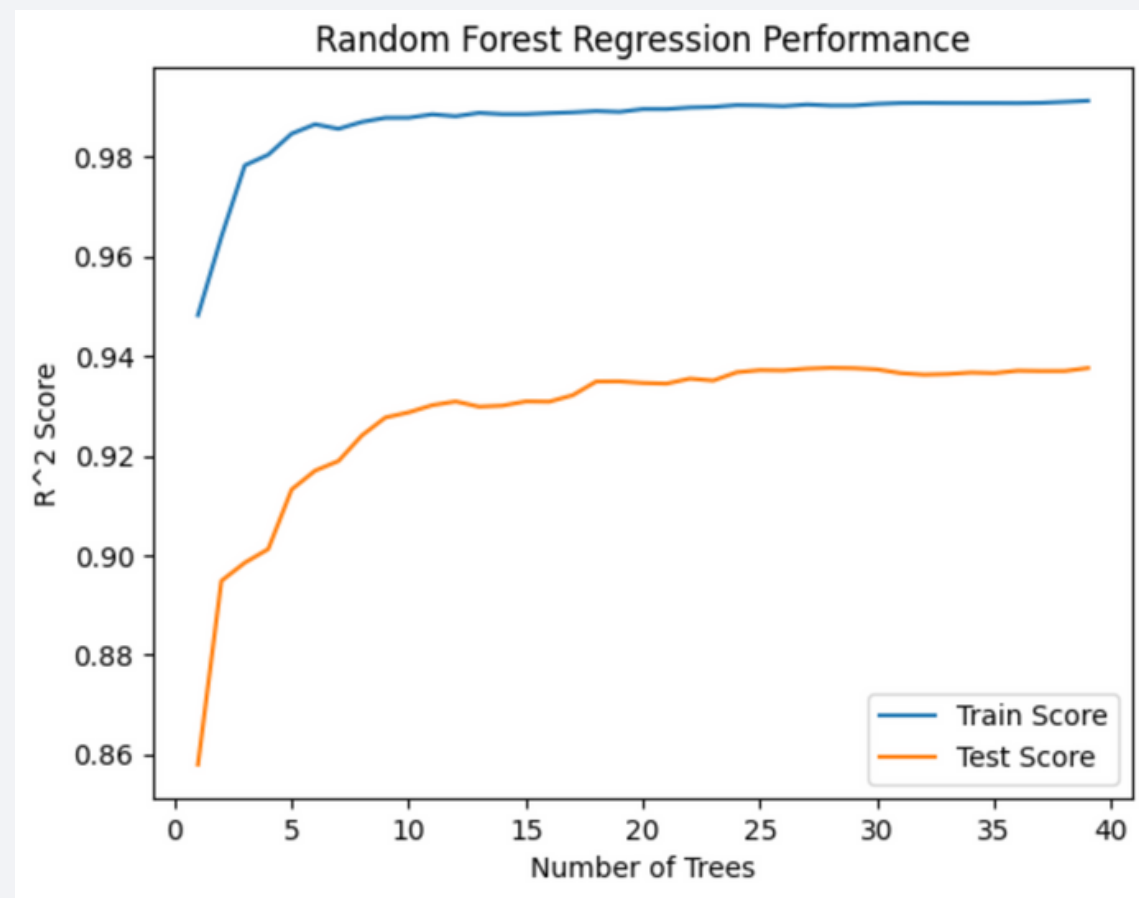
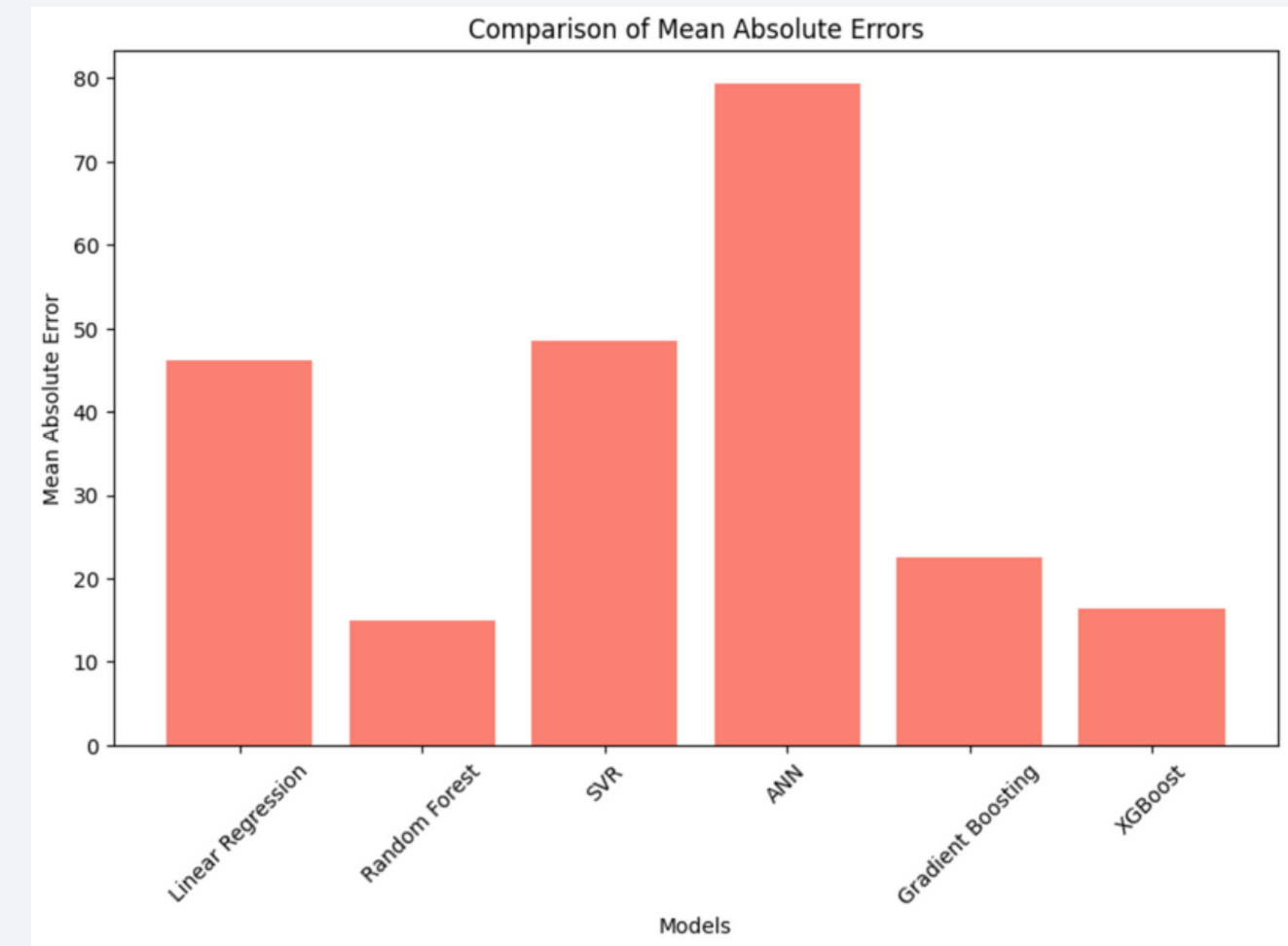
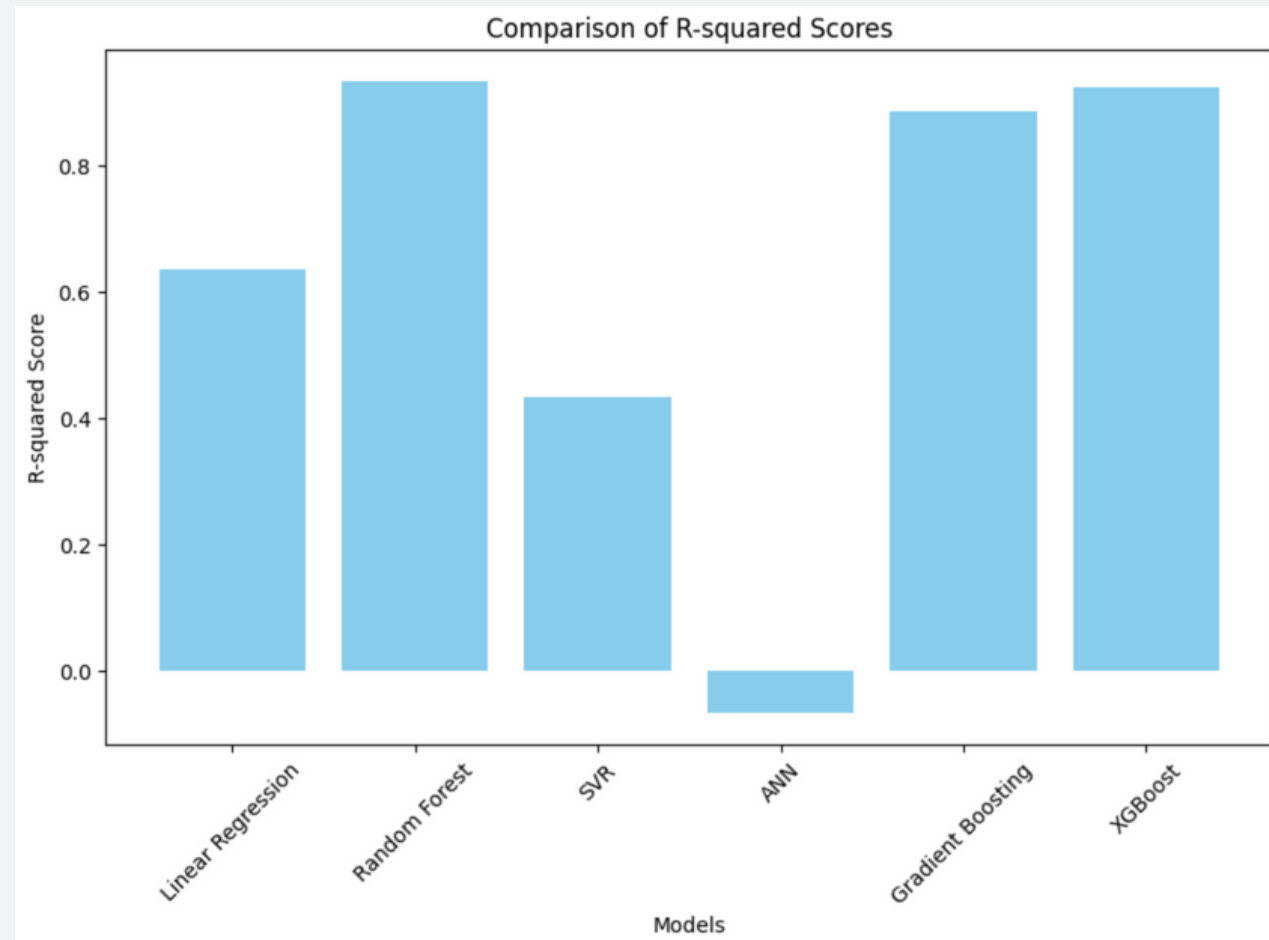


Mean Squared Error: 5608.55
Mean Absolute Error: 48.57
R-squared: 0.43

Mean Squared Error: 1130.51
Mean Absolute Error: 22.46
R-squared: 0.88



Modèle choisi : Random Forest

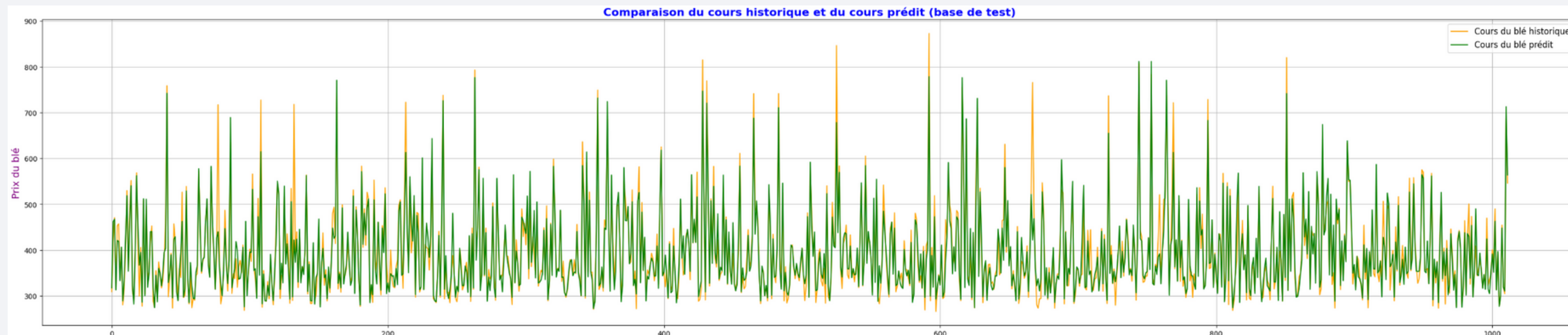


Résultats du projet

Cross-validation du modèle

```
Cross-Validation Scores: [0.69518277 0.7858873 0.71215125 0.6909322 0.68485854 0.75241696  
0.71877689 0.68552867 0.72841676 0.65979666]  
Mean CV Score: 0.7113947990822403
```

Evolution temporelle de notre prédiction vs. historique



Détermination d'un intervalle de confiance de prédiction

```
Valeur historique du cours au 12/04/2024: 387.28  
Intervalle de confiance à 95% de notre prédiction le 12/04/2024: [ 566.52958744 ] ; [ 568.03191256 ] ]
```

Limites du projet

- Pas assez de variables explicatives, en particulier sur les commodities ou les données sont mensualisés tandis que le cours est journalier (exemple: météo, sécheresse..) qui auraient pu être des variables explicatives.
- Très compliqué de vouloir prédire le cours d'une matière première car elle dépend d'énormément de facteurs.
- Data composé de 2500 lignes (2500 dates différentes sur 10 ans)
- Cela aurait été mieux d'avoir plus de data sur lesquels faire travailler notre algorithme mais la data évolue beaucoup d'années en années. Ce qui pose des problèmes car les comportements de marchés d'il y a 20 ans ne sont plus les mêmes aujourd'hui (en particulier sur les mouvements de marchés).